

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> :	A1	(11) International Publication Number: WO 99/16884
C12N 15/49, 15/62, C07K 14/16, A61K 39/21		(43) International Publication Date: 8 April 1999 (08.04.99)
(21) International Application Number: PCT/EP98/06040		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
(22) International Filing Date: 17 September 1998 (17.09.98)		
(30) Priority Data: 9720585.0 26 September 1997 (26.09.97) GB		
(71) Applicant (for all designated States except US): SMITHKLINE BEECHAM BIOLOGICALS S.A. [BE/BE]; Rue de l'Institut 89, B-1330 Rixensart (BE).		
(72) Inventors; and		Published
(75) Inventors/Applicants (for US only): BRUCK, Claudine [BE/BE]; SmithKline Beecham Biologicals S.A., Rue de l'Institut 89, B-1330 Rixensart (BE). GODART, Stephane, Andre, Georges [BE/BE]; SmithKline Beecham Biologicals S.A., Rue de l'Institut 89, B-1330 Rixensart (BE). MARC-HAND, Martine [BE/BE]; SmithKline Beecham Biologicals S.A., Rue de l'Institut 89, B-1330 Rixensart (BE).		With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.
(74) Agent: TYRRELL, Arthur, William, Russell; SmithKline Beecham, Two New Horizons Court, Brentford, Middlesex TW8 9EP (GB).		

(54) Title: FUSION PROTEINS COMPRISING HIV-1 TAT AND/OR NEF PROTEINS

(57) Abstract

The invention provides (a) an HIV Tat protein or derivative thereof linked to either (i) a fusion partner or (ii) an HIV Nef protein or derivative thereof; or (b) an HIV Nef protein or derivative thereof linked to either (i) a fusion partner or (ii) an HIV Tat protein or derivative thereof; or (c) an HIV Nef protein or derivative thereof linked to an HIV Tat protein or derivative thereof and a fusion partner. The invention further provides for a nucleic acid encoding such a protein and a host cell, such as *Pichia Pastoris*, transformed with the aforementioned nucleic acid.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## FUSION PROTEINS COMPRISING HIV-1 TAT AND/OR NEF PROTEINS

The present invention relates to novel HIV protein constructs, to their use in medicine,  
5 to pharmaceutical compositions containing them and to methods of their manufacture.

In particular, the invention relates to fusion proteins comprising HIV-1 Tat and/or Nef  
proteins.

10 HIV-1 is the primary cause of the acquired immune deficiency syndrome (AIDS)  
which is regarded as one of the world's major health problems. Although extensive  
research throughout the world, has been conducted to produce a vaccine, such efforts  
thus far, have not been successful.

15 Non-envelope proteins of HIV-1 have been described and include for example internal  
structural proteins such as the products of the *gag* and *pol* genes and, other non-  
structural proteins such as Rev, Nef, Vif and Tat (Greene et al., New England J. Med,  
324, 5, 308 et seq (1991) and Bryant et al. (Ed. Pizzo), Pediatr. Infect. Dis. J., 11, 5,  
390 et seq (1992).

20 HIV Nef and Tat proteins are early proteins, that is, they are expressed early in  
infection and in the absence of structural proteins.

According to the present invention there is provided a protein comprising  
25 (a) an HIV Nef protein or derivative thereof linked to either (i) a fusion partner or  
(ii) an HIV Tat protein or derivative thereof; or  
(b) an HIV Tat protein or derivative thereof linked to either (i) a fusion partner or  
(ii) an HIV Nef protein or derivative thereof; or  
(c) an HIV Nef protein or derivative thereof linked to an HIV Tat protein or  
30 derivative thereof and a fusion partner.

By 'fusion partner' is meant any protein sequence that is not Tat or Nef.

Preferably the fusion partner is protein D or its' lipidated derivative Lipoprotein D,  
from Haemophilus influenzae B. In particular, it is preferred that the N-terminal

third, i.e. approximately the first 100-130 amino acids are utilised. This is represented herein as Lipo D 1/3. In a preferred embodiment of the invention the Nef protein or derivative thereof may be linked to the Tat protein or derivative thereof. Such Nef-Tat fusions may optionally also be linked to an fusion partner, such as protein D.

5

The fusion partner is normally linked to the N-terminus of the Nef or Tat protein.

Derivatives encompassed within the present invention include molecules with a C terminal Histidine tail which preferably comprises between 5-10 Histidine residues.

- 10 Generally, a histidine tail containing n residues is represented herein as His (n). The presence of an histidine (or 'His') tail aids purification. More specifically, the invention provides proteins with the following structure

	Lipo D 1/3	-	Nef	-	His (6)
15	Lipo D 1/3	-	Nef-Tat	-	His (6)
	Prot D 1/3	-	Nef	-	His (6)
20	Prot D 1/3	-	Nef-Tat	-	His (6)
		Nef-Tat	-	His (6)	

- Figure 1 provides the amino-acid (Seq. ID. No. 7) and DNA sequence (Seq. ID. No. 6) 25 of the fusion partner for such constructs.

In a preferred embodiment the proteins are expressed with a Histidine tail comprising between 5 to 10 and preferably six Histidine residues. These are advantageous in aiding purification. Separate expression, in yeast (*Saccharomyces cerevisiae*), of Nef 30 (Macreadie I.G. et al., 1993, Yeast 9 (6) 565-573) and Tat (Braddock M et al., 1989, Cell 58 (2) 269-79) has already been reported. Nef protein only is myristilated. The present invention provides for the first time the expression of Nef and Tat separately

in a Pichia expression system (Nef-His and Tat-His constructs), and the successful expression of a fusion construct Nef-Tat-His. The DNA and amino acid sequences of representative Nef-His (Seq. ID. No.s 8 and 9), Tat-His (Seq. ID. No.s 10 and 11) and of Nef-Tat-His fusion proteins (Seq. ID. No.s 12 and 13) are set forth in Figure 2.

5

Derivatives encompassed within the present invention also include mutated proteins. The term 'mutated' is used herein to mean a molecule which has undergone deletion, addition or substitution of one or more amino acids using well known techniques for site directed mutagenesis or any other conventional method.

10

A mutated Tat is illustrated in Figure 2 (Seq. ID. No.s 22 and 23) as is a Nef-Tat Mutant-His (Seq. ID. No.s 24 and 25).

15

The present invention also provides a DNA encoding the proteins of the present invention. Such sequences can be inserted into a suitable expression vector and expressed in a suitable host.

A DNA sequence encoding the proteins of the present invention can be synthesized using standard DNA synthesis techniques, such as by enzymatic ligation as described 20 by D.M. Roberts *et al.* in Biochemistry 1985, 24, 5090-5098, by chemical synthesis, by *in vitro* enzymatic polymerization, or by PCR technology utilising for example a heat stable polymerase, or by a combination of these techniques.

Enzymatic polymerisation of DNA may be carried out *in vitro* using a DNA 25 polymerase such as DNA polymerase I (Klenow fragment) in an appropriate buffer containing the nucleoside triphosphates dATP, dCTP, dGTP and dTTP as required at a temperature of 10°-37°C, generally in a volume of 50µl or less. Enzymatic ligation of DNA fragments may be carried out using a DNA ligase such as T4 DNA ligase in an appropriate buffer, such as 0.05M Tris (pH 7.4), 0.01M MgCl<sub>2</sub>, 0.01M 30 dithiothreitol, 1mM spermidine, 1mM ATP and 0.1mg/ml bovine serum albumin, at a temperature of 4°C to ambient, generally in a volume of 50ml or less. The chemical synthesis of the DNA polymer or fragments may be carried out by conventional

phosphotriester, phosphite or phosphoramidite chemistry, using solid phase techniques such as those described in 'Chemical and Enzymatic Synthesis of Gene Fragments - A Laboratory Manual' (ed. H.G. Gassen and A. Lang), Verlag Chemie, Weinheim (1982), or in other scientific publications, for example M.J. Gait, H.W.D. Matthes, M. Singh, B.S. Sproat, and R.C. Titmas, Nucleic Acids Research, 1982, 10, 6243; B.S. Sproat, and W. Bannwarth, Tetrahedron Letters, 1983, 24, 5771; M.D. Matteucci and M.H. Caruthers, Tetrahedron Letters, 1980, 21, 719; M.D. Matteucci and M.H. Caruthers, Journal of the American Chemical Society, 1981, 103, 3185; S.P. Adams *et al.*, Journal of the American Chemical Society, 1983, 105, 661; N.D. Sinha, J. Biernat, J. McMannus, and H. Koester, Nucleic Acids Research, 1984, 12, 4539; and H.W.D. Matthes *et al.*, EMBO Journal, 1984, 3, 801.

The invention also provides a process for preparing a protein of the invention, the  
15 process comprising the steps of :

- i) preparing a replicable or integrating expression vector capable, in a host cell, of expressing a DNA polymer comprising a nucleotide sequence that encodes the protein or  
20 a derivative thereof
- ii) transforming a host cell with said vector
- iii) culturing said transformed host cell under conditions permitting expression of said DNA polymer to produce said protein; and
- iv) recovering said protein  
25

The process of the invention may be performed by conventional recombinant techniques such as described in Maniatis *et al.*, Molecular Cloning - A Laboratory Manual; Cold Spring Harbor, 1982-1989.

30

The term 'transforming' is used herein to mean the introduction of foreign DNA into a host cell. This can be achieved for example by transformation, transfection or

infection with an appropriate plasmid or viral vector using e.g. conventional techniques as described in Genetic Engineering; Eds. S.M. Kingsman and A.J. Kingsman; Blackwell Scientific Publications; Oxford, England, 1988. The term 'transformed' or 'transformant' will hereafter apply to the resulting host cell  
5 containing and expressing the foreign gene of interest.

The expression vectors are novel and also form part of the invention.

The replicable expression vectors may be prepared in accordance with the invention,  
10 by cleaving a vector compatible with the host cell to provide a linear DNA segment having an intact replicon, and combining said linear segment with one or more DNA molecules which, together with said linear segment encode the desired product, such as the DNA polymer encoding the protein of the invention, or derivative thereof, under ligating conditions.

15

Thus, the DNA polymer may be preformed or formed during the construction of the vector, as desired.

The choice of vector will be determined in part by the host cell, which may be  
20 prokaryotic or eukaryotic but preferably is *E. coli* or yeast. Suitable vectors include plasmids, bacteriophages, cosmids and recombinant viruses.

The preparation of the replicable expression vector may be carried out conventionally with appropriate enzymes for restriction, polymerisation and ligation of the DNA, by  
25 procedures described in, for example, Maniatis *et al.* cited above.

The recombinant host cell is prepared, in accordance with the invention, by transforming a host cell with a replicable expression vector of the invention under transforming conditions. Suitable transforming conditions are conventional and are  
30 described in, for example, Maniatis *et al.* cited above, or "DNA Cloning" Vol. II, D.M. Glover ed., IRL Press Ltd, 1985.

- The choice of transforming conditions is determined by the host cell. Thus, a bacterial host such as *E. coli* may be treated with a solution of  $\text{CaCl}_2$  (Cohen *et al.*, Proc. Nat. Acad. Sci., 1973, 69, 2110) or with a solution comprising a mixture of  $\text{RbCl}$ ,  $\text{MnCl}_2$ , potassium acetate and glycerol, and then with 3-[N-morpholino]-  
5 propane-sulphonic acid,  $\text{RbCl}$  and glycerol. Mammalian cells in culture may be transformed by calcium co-precipitation of the vector DNA onto the cells. The invention also extends to a host cell transformed with a replicable expression vector of the invention.
- 10 Culturing the transformed host cell under conditions permitting expression of the DNA polymer is carried out conventionally, as described in, for example, Maniatis *et al.* and "DNA Cloning" cited above. Thus, preferably the cell is supplied with nutrient and cultured at a temperature below 50°C.
- 15 The product is recovered by conventional methods according to the host cell. Thus, where the host cell is bacterial, such as *E. coli* - or yeast such as *Pichia*; it may be lysed physically, chemically or enzymatically and the protein product isolated from the resulting lysate. Where the host cell is mammalian, the product may generally be isolated from the nutrient medium or from cell free extracts. Conventional protein  
20 isolation techniques include selective precipitation, adsorption chromatography, and affinity chromatography including a monoclonal antibody affinity column.

For proteins of the present invention provided with Histidine tails, purification can easily be achieved by the use of a metal ion affinity column. In a preferred embodiment, the protein is further purified by subjecting it to cation ion exchange chromatography and/or Gel filtration chromatography. The protein is then sterilised by passing through a 0.22  $\mu\text{m}$  membrane.

30 The proteins of the invention can then be formulated as a vaccine, or the Histidine residues enzymatically cleared.

The proteins of the present invention are provided preferably at least 80% pure more preferably 90% pure as visualised by SDS PAGE. Preferably the proteins appear as a single band by SDS PAGE.

- 5 The present invention also provides pharmaceutical composition comprising a protein of the present invention in a pharmaceutically acceptable excipient.

Vaccine preparation is generally described in **New Trends and Developments in Vaccines**, Voller *et al.* (eds.), University Park Press, Baltimore, Maryland, 1978.

- 10 Encapsulation within liposomes is described by Fullerton, US Patent 4,235,877.

The proteins of the present invention are preferably adjuvanted in the vaccine formulation of the invention. Suitable adjuvants include an aluminium salt such as aluminium hydroxide gel (alum) or aluminium phosphate, but may also be a salt of calcium, iron or zinc, or may be an insoluble suspension of acylated tyrosine, or acylated sugars, cationically or anionically derivatised polysaccharides, or polyphosphazenes.

- 15 In the formulation of the inventions it is preferred that the adjuvant composition induces a preferential TH1 response. Suitable adjuvant systems include, for example, a combination of monophosphoryl lipid A or derivative thereof, preferably 3-de-O-acylated monophosphoryl lipid A (3D-MPL) together with an aluminium salt.

- 20 An enhanced system involves the combination of a monophosphoryl lipid A and a saponin derivative particularly the combination of QS21 and 3D- MPL as disclosed in WO 94/00153, or a less reactogenic composition where the QS21 is quenched with cholesterol as disclosed in WO 96/33739.

- 25 A particularly potent adjuvant formulation involving QS21, 3D-MPL & tocopherol in an oil in water emulsion is described in WO 95/17210 and is a preferred formulation.

Accordingly in one embodiment of the present invention there is provided a vaccine comprising a protein according to the invention adjuvanted with a monophosphoryl lipid A or derivative thereof, especially 3D-MPL.

- 5 Preferably the vaccine additionally comprises a saponin, more preferably QS21.

Preferably the formulation additional comprises an oil in water emulsion and tocopherol. The present invention also provides a method for producing a vaccine formulation comprising mixing a protein of the present invention together with a 10 pharmaceutically acceptable excipient, such as 3D-MPL.

The vaccine of the present invention may additional comprise further HIV proteins, such as the envelope glycoprotein gp160 or its derivative gp 120.

- 15 In another aspect, the invention relates to an HIV Nef or an HIV Tat protein or derivative thereof expressed in *Pichia pastoris*.

The invention will be further described by reference to the following examples:

20 **EXAMPLES:**

**General**

- 25 Nef and Tat proteins, two regulatory proteins encoded by the human immunodeficiency virus (HIV-1) were produced in *E. coli* and in the methylotrophic yeast *Pichia pastoris*.

The *nef* gene from the Bru/Lai isolate (Cell 40: 9-17, 1985) was selected for these constructs since this gene is among those that are most closely related to the 30 consensus Nef .

The starting material for the Bru/Lai *nef* gene was a 1170bp DNA fragment cloned on the mammalian expression vector pcDNA3 (pcDNA3/*nef*).

The *tat* gene originates from the BH10 molecular clone. This gene was received as an

5 HTLV III cDNA clone named pCV1 and described in Science, 229, p69-73, 1985.

#### 1. EXPRESSION OF HIV-1 *nef* AND *tat* SEQUENCES IN E.COLI.

Sequences encoding the Nef protein as well as a fusion of *nef* and *tat* sequences were

10 placed in plasmids vectors: pRIT14586 and pRIT14589 (see figure 1).

Nef and the Nef-Tat fusion were produced as fusion proteins using as fusion partner a part of the protein D. Protein D is an immunoglobulin D binding protein exposed at the surface of the gram-negative bacterium *Haemophilus influenzae*.

15

pRIT14586 contains, under the control of a λPL promoter, a DNA sequence derived from the bacterium *Haemophilus influenzae* which codes for the first 127 amino acids of the protein D (Infect. Immun. 60 : 1336-1342, 1992), immediately followed by a multiple cloning site region plus a DNA sequence coding for one glycine, 6 histidines 20 residues and a stop codon (Fig. 1A).

This vector is designed to express a processed lipidated His tailed fusion protein (LipoD fusion protein). The fusion protein is synthesised as a precursor with an 18 amino acid residues long signal sequence and after processing, the cysteine at position 25 19 in the precursor molecule becomes the amino terminal residue which is then modified by covalently bound fatty acids (Fig.1B).

pRIT14589 is almost identical to pRIT14586 except that the protD derived sequence starts immediately after the cysteine19 codon.

30 Expression from this vector results in a His tailed, non lipidated fusion protein (Prot D fusion protein).

Four constructs were made: LipoD-*nef*-His, LipoD-*nef-tat*-His, ProtD-*nef*-His, and ProtD-*nef-tat*-His.

The first two constructs were made using the expression vector pRIT14586, the last  
5 two constructs used pRIT14589.

### 1.1 CONSTRUCTION OF THE RECOMBINANT STRAIN ECLD-N1 PRODUCING THE LIPOD-NEF-HIS FUSION PROTEIN.

10

#### 1.1.1 Construction of the lipoD-*nef*-His expression plasmid pRIT14595

The *nef* gene(Bru/Lai isolate) was amplified by PCR from pcDNA3/Nef plasmid with primers 01 and 02.

15

NcoI

PRIMER 01 (Seq ID NO 1): 5' ATCGTCCATG.GGT.GGC.AAG.TGG.T 3'

20

SpeI

PRIMER 02 (Seq ID NO 2): 5' CGGCTACTAGTGCAGTTCTTGAA 3'

The *nef*DNA region amplified starts at nucleotide 8357 and terminates at nucleotide 8971 (Cell, 40: 9-17, 1985).

25

An NcoI restriction site ( which carries the ATG codon of the *nef* gene) was introduced at the 5'end of the PCR fragment while a SpeI site was introduced at the 3' end.

30

The PCR fragment obtained and the expression plasmid pRIT14586 were both restricted by NcoI and SpeI, purified on an agarose gel, ligated and transformed in the

appropriate *E.coli* host cell, strain AR58. This strain is a cryptic  $\lambda$  lysogen derived from N99 that is *galE::Tn10*,  $\Delta$ -8 (*chlD-pgl*),  $\Delta$ -H1 (*cro-chlA*), N<sup>+</sup>, and cI857.

- 5 The resulting recombinant plasmid received, after verification of the *nef* amplified region by automatic sequencing,(see section 1.1.2 below) the pRIT14595 denomination.

### 1.1.2 Selection of transformants of *E. Coli* strain AR58 with pRIT14595.

10

When transformed in AR58 *E.coli* host strain, the recombinant plasmid directs the heat-inducible production of the heterologous protein.

- 15 Heat inducible protein production of several recombinant lipoD-Nef-His transformants was analysed by Coomassie Blue stained SDS-PAGE. All the transformants analysed showed an heat inducible heterologous protein production. The abundance of the recombinant Lipo D-Nef-Tat-His fusion protein was estimated at 10% of total protein.
- 20 One of the transformants was selected and given the laboratory accession number ECLD-N1.

- 25 The recombinant plasmid was reisolated from strain ECLD-N1, and the sequence of the *nef*-His coding region was confirmed by automated sequencing .This plasmid received the official designation pRIT14595.

The fully processed and acylated recombinant Lipo D-*nef*-His fusion protein produced by strain ECLD-N1 is composed of:

- 30 °Fatty acids  
°109 a.a. of proteinD (starting at a.a.19 and extending to a.a.127).

°A methionine, created by the use of NcoI cloning site of pRIT14586  
(Fig.1).

°205a.a. of Nef protein (starting at a.a.2 and extending to a.a.206).

5 °A threonine and a serine created by the cloning procedure (cloning at SpeI site of pRIT14586).

°One glycine and six histidines.

## 1.2 CONSTRUCTION OF RECOMBINANT STRAIN ECD-N1 PRODUCING PROT D-Nef-HIS FUSION PROTEIN.

10

Construction of expression plasmid pRIT14600 encoding the Prot D-Nef-His fusion protein was identical to the plasmid construction described in example 1.1.1 with the exception that pRIT14589 was used as receptor plasmid for the PCR amplified *nef* fragment.

15

E.coli AR58 strain was transformed with pRIT14600 and transformants were analysed as described in example 1.1.2. The transformant selected received laboratory accession number ECD-N1.

**1.3 CONSTRUCTION OF RECOMBINANT STRAIN ECLD-NT6  
PRODUCING THE LIPO D-Nef-Tat-HIS FUSION PROTEIN.**

**1.3.1 Construction of the lipo D-Nef-Tat-His expression plasmid pRIT14596**

5

The *tat* gene(BH10 isolate) was amplified by PCR from a derivative of the pCV1 plasmid with primers 03 and 04. SpeI restriction sites were introduced at both ends of the PCR fragment.

10

SpeI

PRIMER 03 (Seq ID NO 3): 5' ATCGTACTAGT.GAG.CCA.GTA.GAT.C 3'

SpeI

PRIMER 04 (Seq ID NO 4): 5' CGGCTACTAGTTTCCCTCGGGCCT 3'

15

The nucleotide sequence of the amplified *tat* gene is illustrated in the pCV1 clone (Science 229 : 69-73, 1985) and covers nucleotide 5414 till nucleotide 7998.

20 The PCR fragment obtained and the plasmid pRIT14595 (expressing lipoD-Nef-His protein) were both digested by SpeI restriction enzyme, purified on an agarose gel, ligated and transformed in competent AR58 cells. The resulting recombinant plasmid received, after verification of the *tat* amplified sequence by automatic sequencing (see section 1.3.2 below), the pRIT14596 denomination.

25

**1.3.2 Selection of transformants of strain AR58 with pRIT14596**

Transformants were grown, heat induced and their proteins were analysed by Coomassie Blue stained gels. The production level of the recombinant protein was 30 estimated at 1% of total protein. One recombinant strain was selected and received the laboratory denomination ECLD-NT6.

The lipoD-*nef-tat*-His recombinant plasmid was reisolated from ECLD-NT6 strain, sequenced and received the official designation pRIT14596.

The fully processed and acylated recombinant Lipo D-Nef-Tat-His fusion protein

5 produced by strain ECLD-N6 is composed of:

°Fatty acids

°109 a.a. of proteinD (starting at a.a.19 and extending to a.a.127).

°A methionine, created by the use of NcoI cloning site of pRIT14586.

10 °205a.a. of the Nef protein (starting at a.a.2 and extending to a.a.206)

°A threonine and a serine created by the cloning procedure

°85a.a. of the Tat protein (starting at a.a.2 and extending to a.a.86)

°A threonine and a serine introduced by cloning procedure

°One glycine and six histidines.

15

#### **1.4 CONSTRUCTION OF RECOMBINANT STRAIN ECD-NT1 PRODUCING PROT D-Nef-Tat-HIS FUSION PROTEIN.**

Construction of expression plasmid pRIT14601 encoding the Prot D-Nef-Tat-His

20 fusion protein was identical to the plasmid construction described in example 1.3.1  
with the exception that pRIT14600 was used as receptor plasmid for the PCR  
amplified *nef* fragment.

*E.coli* AR58 strain was transformed with pRIT14601 and transformants were analysed  
25 as described previously. The transformant selected received laboratory accession  
number ECD-NT1.

30

2. EXPRESSION OF HIV-1 *nef* AND *tat* SEQUENCES IN PICHIA PASTORIS.

Nef protein, Tat protein and the fusion Nef-Tat were expressed in the methylotrophic yeast *Pichia pastoris* under the control of the inducible alcohol oxidase (AOX1) promoter.

To express these HIV-1 genes a modified version of the integrative vector PHIL-D2 (INVITROGEN) was used. This vector was modified in such a way that expression of heterologous protein starts immediately after the native ATG codon of the AOX1 gene and will produce recombinant protein with a tail of one glycine and six histidines residues. This PHIL-D2-MOD vector was constructed by cloning an oligonucleotide linker between the adjacent AsuII and EcoRI sites of PHIL-D2 vector (see Figure 3). In addition to the His tail, this linker carries NcoI, SpeI and XbaI restriction sites between which *nef*, *tat* and *nef-tat* fusion were inserted.

2.1 CONSTRUCTION OF THE INTEGRATIVE VECTORS pRIT14597 (encoding Nef-His protein), pRIT14598 (encoding Tat-His protein) and pRIT14599 (encoding fusion Nef-Tat-His).

The *nef* gene was amplified by PCR from the pcDNA3/Nef plasmid with primers 01 and 02 (see section 1.1.1 construction of pRIT14595). The PCR fragment obtained and the integrative PHIL-D2-MOD vector were both restricted by NcoI and SpeI, purified on agarose gel and ligated to create the integrative plasmid pRIT14597 (see Figure 3).

The *tat* gene was amplified by PCR from a derivative of the pCV1 plasmid with primers 05 and 04 (see section 1.3.1 construction of pRIT14596):

NcoI

30 PRIMER 05 (Seq ID NO 5): 5'ATCGTCCATGGAGCCAGTAGATC 3'

An NcoI restriction site was introduced at the 5' end of the PCR fragment while a SpeI site was introduced at the 3' end with primer 04. The PCR fragment obtained and the PHIL-D2-MOD vector were both restricted by NcoI and SpeI, purified on agarose gel and ligated to create the integrative plasmid pRIT14598.

5

To construct pRIT14599, a 910bp DNA fragment corresponding to the *nef-tat*-His coding sequence was ligated between the EcoRI blunted(T4 polymerase) and NcoI sites of the PHIL-D2-MOD vector. The *nef-tat*-His coding fragment was obtained by XbaI blunted(T4 polymerase) and NcoI digestions of pRIT14596.

10

## 2.2 TRANSFORMATION OF PICHIA PASTORIS STRAIN GS115(his4).

To obtain *Pichia pastoris* strains expressing Nef-His, Tat-His and the fusion Nef-Tat-His, strain GS115 was transformed with linear NotI fragments carrying the respective expression cassettes plus the HIS4 gene to complement his4 in the host genome. Transformation of GS115 with NotI-linear fragments favors recombination at the AOX1 locus.

Multicopy integrant clones were selected by quantitative dot blot analysis and the type of integration, insertion (Mut<sup>+</sup>phenotype) or transplacement (Mut<sup>\*</sup>phenotype), was determined.

From each transformation, one transformant showing a high production level for the recombinant protein was selected :

25

Strain Y1738 (Mut<sup>+</sup> phenotype) producing the recombinant Nef-His protein, a myristylated 215 amino acids protein which is composed of:

°Myristic acid

30 °A methionine, created by the use of NcoI cloning site of PHIL-D2-MOD vector

°205 a.a. of Nef protein(starting at a.a.2 and extending to a.a.206)

- °A threonine and a serine created by the cloning procedure (cloning at SpeI site of PHIL-D2-MOD vector.
- °One glycine and six histidines.

5 Strain Y1739 (Mut<sup>+</sup> phenotype) producing the Tat-His protein, a 95 amino acid protein which is composed of:

- °A methionine created by the use of NcoI cloning site
- °85 a.a. of the Tat protein(starting at a.a.2 and extending to a.a.86)

10

- °A threonine and a serine introduced by cloning procedure
- °One glycine and six histidines

Strain Y1737(Mut<sup>+</sup> phenotype) producing the recombinant Nef-Tat-His fusion protein,  
15 a myristylated 302 amino acids protein which is composed of:

- °Myristic acid
  - °A methionine, created by the use of NcoI cloning site
  - °205a.a. of Nef protein(starting at a.a.2 and extending to a.a.206)
- 20
- °A threonine and a serine created by the cloning procedure
  - °85a.a. of the Tat protein(starting at a.a.2 and extending to a.a.86)
  - °A threonine and a serine introduced by the cloning procedure
  - °One glycine and six histidines

**3. EXPRESSION OF HIV-1 Tat-MUTANT IN PICHIA PASTÓRIS**

As well as a Nef-Tat mutant fusion protein, a mutant recombinant Tat protein has also  
5 been expressed. The mutant Tat protein must be biologically inactive while  
maintaining its immunogenic epitopes.

A double mutant *tat* gene, constructed by D.Clements (Tulane University) was  
selected for these constructs.

10

This *tat* gene (originates from BH10 molecular clone) bears mutations in the active site region (Lys41→Ala) and in RGD motif (Arg78→Lys and Asp80→Glu) (*Virology* 235: 48-64, 1997).

15 The mutant *tat* gene was received as a cDNA fragment subcloned between the EcoRI and HindIII sites within a CMV expression plasmid (pCMVLys41/KGE)

**3.1 CONSTRUCTION OF THE INTEGRATIVE VECTORS**

20 pRIT14912(encoding Tat mutant-His protein) and pRIT14913(encoding fusion Nef-Tat mutant-His).

The *tat* mutant gene was amplified by PCR from the pCMVLys41/KGE plasmid with primers 05 and 04 (see section 2.1 construction of pRIT14598)

25

An NcoI restriction site was introduced at the 5' end of the PCR fragment while a Spel site was introduced at the 3' end with primer 04. The PCR fragment obtained and the PHIL-D2-MOD vector were both restricted by NcoI and Spel, purified on agarose gel and ligated to create the integrative plasmid pRIT14912

30

To construct pRIT14913, the *tat* mutant gene was amplified by PCR from the pCMVLys41/KGE plasmid with primers 03 and 04 (see section 1.3.1 construction of pRIT14596).

- 5 The PCR fragment obtained and the plasmid pRIT14597 (expressing Nef-His protein) were both digested by SpeI restriction enzyme, purified on agarose gel and ligated to create the integrative plasmid pRIT14913

### **3.2 TRANSFORMATION OF PICHIA PASTORIS STRAIN GS115.**

10

Pichia pastoris strains expressing Tat mutant-His protein and the fusion Nef-Tat mutant-His were obtained, by applying integration and recombinant strain selection strategies previously described in section 2.2 .

- 15 Two recombinant strains producing Tat mutant-His protein ,a 95 amino-acids protein, were selected: Y1775 (Mut<sup>+</sup> phenotype) and Y1776(Mut<sup>+</sup> phenotype).

One recombinant strain expressing Nef-Tat mutant-His fusion protein, a 302 amino-acids protein was selected: Y1774(Mut<sup>+</sup> phenotype).

20

**4. PURIFICATION OF Nef-Tat-His FUSION PROTEIN (PICHIA PASTORIS)**

- 5 The purification scheme has been developed from 146g of recombinant Pichia pastoris cells (wet weight) or 2L Dyno-mill homogenate OD 55. The chromatographic steps are performed at room temperature. Between steps , Nef-Tat positive fractions are kept overnight in the cold room (+4°C) ; for longer time, samples are frozen at -20°C.

10

146g of Pichia pastoris cells



Homogenization

Buffer: 2L 50 mM PO<sub>4</sub>, pH 7.0

final OD:50



Dyno-mill disruption (4 passes)



Centrifugation

JA10 rotor / 9500 rpm/ 30 min / room  
temperature

Dyno-mill Pellet



Wash

Buffer: +2L 10 mM PO<sub>4</sub>, pH 7.5 -  
150mM - NaCl 0,5% empigen

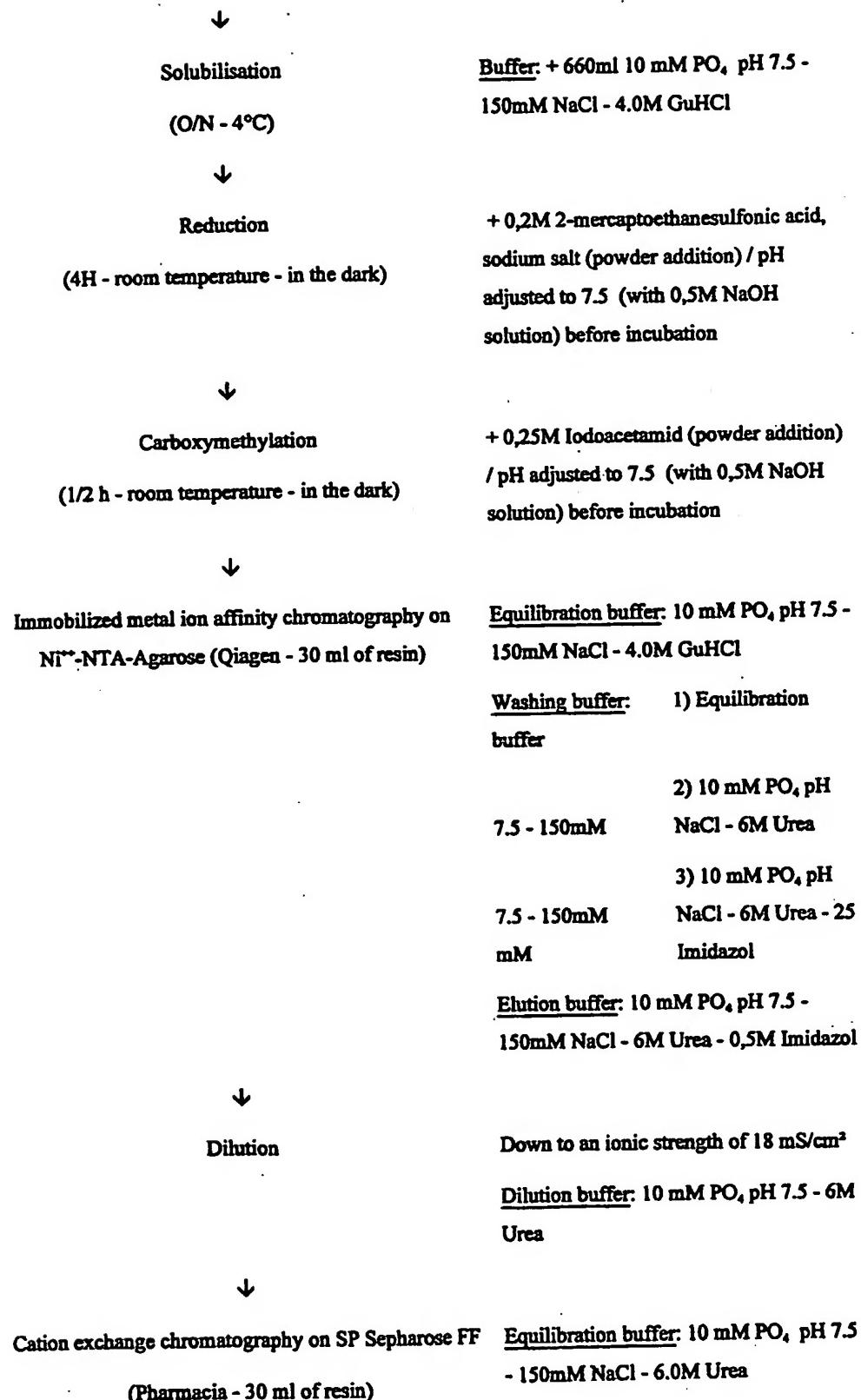
(1h - 4°C)



Centrifugation

JA10 rotor / 9500 rpm/ 30 min / room  
temperature

Pellet



	<u>Washing buffer:</u>	1) Equilibration buffer
		2) 10 mM PO <sub>4</sub> pH 7.5 - 250mM NaCl - 6M Urea
	<u>Elution buffer:</u>	10 mM Borate pH 9.0 - 2M NaCl - 6M Urea
↓		
Concentration		up to 5 mg/ml 10kDa Omega membrane(Filtron)
		↓
Gel filtration chromatography on Superdex200 XK 16/60 (Pharmacia - 120 ml of resin)		<u>Elution buffer:</u> 10 mM PO <sub>4</sub> pH 7.5 - 150mM NaCl - 6M Urea 5 ml of sample / injection → 5 injections
		↓
Dialysis (O/N - 4°C)		<u>Buffer:</u> 10 mM PO <sub>4</sub> pH 6.8 - 150mM NaCl - 0,5M Arginin*
		↓
Sterile filtration		Millex GV 0,22μm

\* ratio: 0,5M Arginin for a protein concentration of 1600μg/ml.

## 5 Purity

The level of purity as estimated by SDS-PAGE is shown in Figure 4 by Daiichi Silver Staining and in Figure 5 by Coomassie blue G250.

After Superdex200 step: > 95%

After dialysis and sterile filtration steps: > 95%

## 5 Recovery

51mg of Nef-Tat-his protein are purified from 146g of recombinant Pichia pastoris cells (= 2L of Dyno-mill homogenate OD 55)

## 10 5. VACCINE PREPARATION

A vaccine prepared in accordance with the invention comprises the expression product of a DNA recombinant encoding an antigen as exemplified in example 1 or 2 and as adjuvant, the formulation comprising a mixture of 3 de -O-acylated monophosphoryl

15 lipid A 3D-MPL and QS21 in an oil/water emulsion.

**3D-MPL:** is a chemically detoxified form of the lipopolysaccharide (LPS) of the Gram-negative bacteria *Salmonella minnesota*.

20 Experiments performed at Smith Kline Beecham Biologicals have shown that 3D-MPL combined with various vehicles strongly enhances both the humoral and a TH1 type of cellular immunity.

25 **QS21:** is one saponin purified from a crude extract of the bark of the Quillaja Saponaria Molina tree, which has a strong adjuvant activity: it activates both antigen-specific lymphoproliferation and CTLs to several antigens. Experiments performed at Smith Kline Beecham Biologicals have demonstrated a clear synergistic effect of combinations of 3D-MPL and QS21 in the induction of both humoral and TH1 type cellular immune responses.

30

The oil/water emulsion is composed of 2 oils (a tocopherol and squalene), and of PBS containing Tween 80 as emulsifier. The emulsion comprised 5% squalene 5%

tocopherol 0.4% Tween 80 and had an average particle size of 180 nm (see WO 95/17210).

Experiments performed at Smith Kline Beecham Biologicals have proven that the  
5 adjunction of this O/W emulsion to 3D-MPL/QS21 further increases their immunostimulant properties.

**Preparation of the oil/water emulsion (2 fold concentrate)**

10 Tween 80 is dissolved in phosphate buffered saline (PBS) to give a 2% solution in the PBS. To provide 100ml two fold concentrate emulsion 5g of DL alpha tocopherol and 5ml of squalene are vortexed to mix thoroughly. 90ml of PBS/Tween solution is added and mixed thoroughly. The resulting emulsion is then passed through a syringe and finally microfluidised by using an M110S microfluidics machine. The resulting  
15 oil droplets have a size of approximately 180 nm.

**Preparation of oil in water formulation.**

Antigen prepared in accordance with example 1 or 2 (5 $\mu$ g) was diluted in 10 fold  
20 concentrated PBS pH 6.8 and H<sub>2</sub>O before consecutive addition of SB62, 3D-MPL (5 $\mu$ g), QS21 (5 $\mu$ g) and 50  $\mu$ g/ml thiomersal as preservative at 5 min interval. The emulsion volume is equal to 50% of the total volume (50 $\mu$ l for a dose of 100 $\mu$ l).

All incubations were carried out at room temperature with agitation.

25

**6. IMMUNOGENICITY OF Tat AND Nef-Tat IN RODENTS**

Characterization of the immune response induced after immunization with Tat and  
30 NefTat was carried out. To obtain information on isotype profiles and cell-mediated immunity (CMI) two immunization experiments in mice were conducted. In the first experiment mice were immunized twice two weeks apart into the footpad with Tat or

- NefTat in the oxydized or reduced form, respectively. Antigens were formulated in an oil in water emulsion comprising squalene, tween 80 <sup>TM</sup> (polyoxyethylene sorbitan monooleate) QS21, 3D-MPL and  $\alpha$ -tocopherol, and a control group received the adjuvant alone. Two weeks after the last immunization sera were obtained and
- 5 subjected to Tat-specific ELISA (using reduced Tat for coating) for the determination of antibody titers and isotypes (Figure 6a). The antibody titers were highest in the mice having received oxydized Tat. In general, the oxydized molecules induced higher antibody titers than the reduced forms, and Tat alone induced higher antibody titers than NefTat. The latter observation was confirmed in the second experiment.
- 10 Most interestingly, the isotype profile of Tat-specific antibodies differed depending on the antigens used for immunization. Tat alone elicited a balanced IgG1 and IgG2a profile, while NefTat induced a much stronger T<sub>H2</sub> bias (Figure 6b). This was again confirmed in the second experiment.
- 15 In the second mouse experiment animals received only the reduced forms of the molecules or the adjuvant alone. Besides serological analysis (see above) lymphoproliferative responses from lymph node cells were evaluated. After restimulation of those cells in vitro with Tat or NefTat <sup>3</sup>H-thymidine incorporation was measured after 4 days of culture. Presentation of the results as stimulation indices
- 20 indicates that very strong responses were induced in both groups of mice having received antigen (Figure 7).

In conclusion, the mice studies indicate that Tat as well as Nef-Tat are highly immunogenic candidate vaccine antigens. The immune response directed against the

25 two molecules is characterized by high antibody responses with at least 50% IgG1. Furthermore, strong CMI responses (as measured by lymphoproliferation) were observed.

## 7. FUNCTIONAL PROPERTIES OF THE Tat AND Nef-Tat PROTEINS

30

The Tat and NefTat molecules in oxydized or reduced form were investigated for their ability to bind to human T cell lines. Furthermore, the effect on growth of

those cell lines was assessed. ELISA plates were coated overnight with different concentration of the Tat and NefTat proteins, the irrelevant gD from herpes simplex virus type II, or with a buffer control alone. After removal of the coating solution HUT-78 cells were added to the wells. After two hours of incubation the wells were 5 washed and binding of cells to the bottom of the wells was assessed microscopically. As a quantitative measure cells were stained with toluidine blue, lysed by SDS, and the toluidine blue concentration in the supernatant was determined with an ELISA plate reader. The results indicate that all four proteins, Tat and NefTat in oxydized or reduced form mediated binding of the cells to the 10 ELISA plate (Figure 8). The irrelevant protein (data not shown) and the buffer did not fix the cells. This indicates that the recombinantly expressed Tat-containing proteins bind specifically to human T cell lines.

In a second experiment HUT-78 cells were left in contact with the proteins for 16 15 hours. At the end of the incubation period the cells were labeled with [<sup>3</sup>H]-thymidine and the incorporation rate was determined as a measure of cell growth. All four proteins included in this assay inhibited cell growth as judged by diminished radioactivity incorporation (Figure 9). The buffer control did not 20 mediate this effect. These results demonstrate that the recombinant Tat-containing proteins are capable of inhibiting growth of a human T cell line.

In summary the functional characterization of the Tat and NefTat proteins reveals that these proteins are able to bind to human Tcell lines. Furthermore, the proteins are able to inhibit growth of such cell lines.

**CLAIMS**

1. A protein comprising
  - 5 (a) an HIV Tat protein or derivative thereof linked to either (i) a fusion partner or (ii) an HIV Nef protein or derivative thereof; or
  - (b) an HIV Nef protein or derivative thereof linked to either (i) a fusion partner or (ii) an HIV Tat protein or derivative thereof; or
  - (c) an HIV Nef protein or derivative thereof linked to an HIV Tat protein or derivative thereof and a fusion partner.
- 15 2. A protein as claimed in claim 1 which is a Tat-Nef fusion protein or derivative thereof.
3. A protein as claimed in claim 1 which is a Nef-Tat fusion protein or derivative thereof.
- 20 4. A protein according to claim 1 wherein the derivative of the Tat protein is a mutated Tat protein.
5. A protein according to claim 1 wherein the derivative of the Nef protein is a mutated Nef protein.
- 25 6. A Protein as claimed in any one of claims 1 - 5 wherein the fusion partner is a lipoprotein or derivative thereof.
7. A protein as claimed in claim 6 wherein the lipoprotein is Haemophilus Influenza B protein D or derivative thereof.

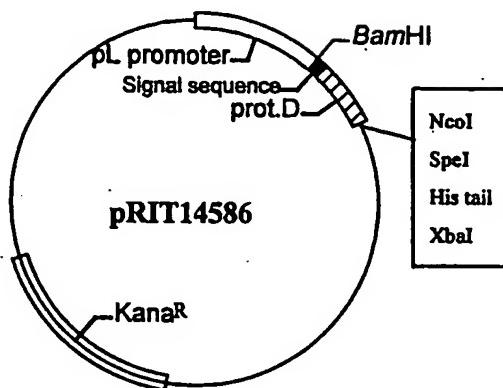
30

8. A protein as claimed in Claim 7 wherein the fusion partner comprises between 100-130 amino acid from the N terminal of Haemophilus Influenza B protein D.
- 5 9. A protein as claimed in any one of Claims 1 to 8, wherein the Tat protein is the entire Tat protein.
10. 10. A protein as claimed in any one of Claims 1 to 8, wherein the Nef protein is the entire Nef protein.
11. 11. A protein as claimed in any one of Claims 1 to 10, wherein the Tat protein is fused to an HIV Nef protein and a fusion partner.
12. 15. A protein as claimed in any one of Claims 1 to 11, wherein the protein has a Histidine tail.
13. 20. A nucleic acid encoding a protein of Claims 1 to 12.
14. 25. A host transformed with a nucleic acid of Claim 13.
15. A host as claimed in claim 14 wherein the host is either Pichia pastoris or E. coli.
16. 25. A vaccine comprising a protein of any one of Claims 1 to 12 in admixture with a pharmaceutically acceptable excipient.
17. 30. A vaccine of Claim 16 additionally comprising an adjuvant.
18. A vaccine of claim 17 wherein the adjuvant is a TH1 inducing adjuvant.

19. A vaccine as claimed in Claim 17 or 18 which adjuvant comprises monophosphoryl lipid A or derivative thereof such as 3 de-O-acylated monophosphoryl lipid A.
- 5 20. A vaccine as claimed in any one of Claims 16 to 19 additionally comprising a saponin adjuvant.
- 10 21. A method of producing a protein of Claim 1 to 12, comprising the steps of transforming a host with a nucleic acid encoding said protein, expressing said protein and recovering the protein.
22. A method as claimed in Claim 21 wherein the host is *E. coli.* or *Pichia pastoris.*
- 15 23. A method of producing a vaccine of Claim 16 to 20, comprising admixing the protein of Claim 1 to 12 with a pharmaceutically acceptable diluent.
24. A method of preparing (i) an HIV Nef protein or derivative thereof or (ii) an HIV Tat protein or derivative thereof in *Pichia pastoris* which method  
20 comprises the steps of transforming *Pichia pastoris* with DNA encoding said HIV Nef protein or derivative thereof or HIV Tat protein or derivative thereof, expressing said protein and recovering the protein.

25

30

**Figure 1:** A/ Map of plasmid pRIT14586

B/ Coding sequence of the first 127 amino acids  
of protein D and multiple cloning site. The signal  
sequence is underlined.

```

BamHI
ATG GAT CCA AAA ACT TTA GCC CTT TCT TTA TTA GCA GCT GCC GTA CTA GCA GGT TGT AGC AGC
Met Asp Pro Lys Thr Leu Ala Leu Ser Leu Leu Ala Ala Gly Val Leu Ala Gly Cys Ser Ser
CAT TCA TCA AAT ATG GCG AAT ACC CAA ATG AAA TCA GAC AAA ATC ATT ATT GCT CAC CGT GGT
His Ser Ser Asn Met Ala Asn Thr Gln Met Lys Ser Asp Lys Ile Ile Ile Ala His Arg Gly
GCT AGC GGT TAT TTA CCA GAG CAT ACG TTA GAA TCT AAA GCA CTT GCT TTT GCA CAA CAG GCT
Ala Ser Gly Tyr Leu Pro Gln His Thr Leu Gln Ser Lys Ala Leu Ala Phe Ala Gln Gln Ala
GAT TAT TTA GAG CAA GAT TTA GCA ATG ACT AAG GAT GGT CGT TTA GTG GTT ATT CAC GAT CAC
Asp Tyr Leu Gln Gln Asp Leu Ala Met Thr Lys Asp Gly Arg Leu Val Val Ile His Asp His
TTT TTA GAT GGC TTG ACT GAT GTT GCG AAA AAA TTC CCA CAT CGT CAT CGT AAA GAT GGC CGT
Phe Leu Asp Gly Leu Thr Asp Val Ala Lys Lys Phe Pro His Arg His Arg Lys Asp Gly Arg
TACT TAT GTC ATC GAC TTT ACC TTA AAA GAA ATT GAA AGT TTA GAA ATG ACA GAA AAC TTT GAA
Tyr Tyr Val Ile Asp Phe Thr Leu Lys Gln Ile Gln Ser Leu Glu Met Thr Glu Asn Phe Gln
Ncol SpeI XbaI
ACC ATG GCC ACG TGT GAT CAG AGC TCA ACT AGT GGA CAC CAT CAC CAT CAC CAT TAA TCT AGA
Thr Met Ala Thr Cys Asp Gln Ser Ser Thr Ser Gly His His His His His His His His His

```

The amino acid sequence of Figure 1 relates to Seq. ID no. 7 and the nucleic acid sequence of Figure 1 relates to Seq. ID. No. 6.

The DNA and amino acid sequences of Nef-His; Tat-His; Nef-Tat-His fusion and mutated Tat is illustrated.

**Pichia-expressed constructs (plain constructs)**

⇒ Nef - HIS

**DNA sequence (Seq. ID. No. 8)**

ATGGGTGGCAAGTGGTCAAAAGTAGTGTGGTTGGATGGCCTACTGTAAGGGAAAGA  
ATGAGACGAGCTGAGCCAGCAGCATGGGTGGGAGCAGCATCTCGAGACCTGGAA  
AAACATGGAGCAATACAAGTAGCAATAACAGCAGCTACCAATGCTGCTTGTGCCTGG  
CTAGAAGCACAAGAGGAGGAGGAGGTGGGTTTCCAGTCACACCTCAGGTACCTTA  
AGACCAATGACTTACAAGGCAGCTGTAGATCTTAGCCACTTTTAAAAGAAAAGGGG  
GGACTGGAAGGGCTAATTCACTCCAACGAAGACAAGATATCCTTGATCTGTGGATC  
TACCACACACAAGGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCAGGGGTC  
AGATATCCACTGACCTTGGATGGTGTACAAGCTAGTACCACTTGAGCAGATAAG  
GTAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTACACCCTGTGAGCCTGCAT  
GGAATGGATGACCCCTGAGAGAGAAAGTGTAGTGGAGGTTGACAGCCGCTAGCA  
TTTCATCACGTGGCCCGAGAGCTGCATCCGGAGTACTTCAAGAACTGCACTAGTGGC  
CACCACCATCACCATTAA

**Protein sequence (Seq. ID. No. 9)**

MGGKWSKSSVVGWPTVRERMRAEPAADGVGAASRDLEKHGAIITSSNTAATNAACAW  
LEAQEEEVGFPTVPQVPLRPMTYKAADVLSHFLKEKGGLIHSQRQDILDLWI  
YHTQGYFPDWQNYTPGPVRYPLTFGWCYKLVPVEPDKVEEANKGENTSLLHPVSLH  
GMDDPEREVLEWRFDSRLAFHVARELHPEYFKNCTSGHzHHHHH.

⇒ Tat - HIS

**DNA sequence (Seq. ID. No. 10)**

ATGGAGCCAGTAGATCCTAGACTAGAGCCCTGGAAGCATCCAGGAAGTCAGCCTAAA  
ACTGCTTGTACCAATTGCTATTGTAAAAAGTGTGCTTCAATTGCCAAGTTGTTTC  
ATAACAAAAGCCTTAGGCATCTCCTATGGCAGGAAGAAGCGGAGACAGCGACGAAGA  
CCTCCTCAAGGCAGTCAGACTCATCAAGTTCTATCAAAGCAACCCACCTCCAA

TCCCGAGGGACCGACAGGCCGAAGGAAACTAGTGGCACCATCACCATCACCAT  
TAA

Protein sequence (Seq. ID. No. 11)

MEPVDPRLPEWKPGSQPKTACTNCYCKCCFHCQVCFITKALGISYGRKKRRQRRR  
PPQGSQTHQVSLSKQPTSQRGDPTGPKETSGHHHHHH.

⇒ Nef - Tat - HIS

DNA sequence (Seq. ID. No. 12)

ATGGGTGGCAAGTGGTCAAAAGTAGTGTGGTTGGATGGCCTACTGTAAGGGAAAGA  
ATGAGACGAGCTGAGCCAGCAGCAGATGGGGTGGGAGCAGCATCTGAGACCTGGAA  
AAACATGGAGCAATCACAAGTAGCAATAACAGCAGCTACCAATGCTGCTTGCGCTGG  
CTAGAACAGAACAGAGGAGGAGGAGGAGGTTTCCAGTCACACCTCAGGTACCTTTA  
AGACCAATGACTTACAAGGCAGCTGTAGATCTTAGCCACTTTTAAAAGAAAAGGGG  
GGACTGGAAGGGCTAATTCACTCCCAACGAAGACAAGATACTCTTGATCTGTGGATC  
TACCACACACAAGGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGTC  
AGATATCCACTGACCTTGGATGGTGTACAAGCTAGTACCAAGCTTGGAGCCAGATAAG  
GTAGAAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTACACCCTGTGAGCCTGCAT  
GGAATGGATGACCTGAGAGAGAAGTGTAGAGTGGAGGTTGACAGCCGCCTAGCA  
TTTCATCACGTGGCCCGAGAGCTGCATCCGGAGTACTCAAGAACTGCACACTAGTGAG  
CCAGTAGATCCTAGACTAGAGCCCTGGAAGCATTCCAGGAAGTCAGCCTAAAATGCT  
TGTACCAATTGCTATTGTAAGGTGTGCTTCAATTGCCAAGTTGTTCTATAACA  
AAAGCCTTAGGCATCTCTATGGCAGGAAGAACGGAGACAGCGACGAAGACCTCCT  
CAAGGCAGTCAGACTCATCAAGTTCTCTATCAAAGCAACCCACCTCCAAATCCCGA  
GGGGACCCGACAGGCCGAAGGAAACTAGTGGCCACCATCACCATCACCATTAAC

Protein sequence (Seq. ID. No. 13)

^  
MGGKWSKSSVVGWPTVRERMRAEPAADGVGAASRDLEKHGAITSNTAATNAACAW  
LEAQEEEVGFVTPQVPLRPMTYKAADVLSHFLKEKGLEGLIHSQRQDILDLWI  
YHTQGYPPDWQNYTPGPVRYPLTFGWCYKLVPVEPDVKVEEANKGENTSLHPVSLH  
GMDDPEREVLEWRFDSSLAFHHVARELHPEYFKNCTSEPVDPRLPEWKPGSQPKTA  
CTNCYCKCCFHCQVCFITKALGISYGRKKRRQRRPPQGSQTHQVSLSKQPTSQR  
GDPTGPKETSGHHHHHH.

E.coli-expressed constructs (fusion constructs)

⇒ LipoD-Nef-HIS

4/17

DNA sequence (Seq. ID. No. 14)

Nucleotides corresponding to the Prot D Fusion Partner are in bold.  
 The Lipidation Signal Sequence is underlined. After processing, the cysteine coded by the TGT codon, indicated with a star, becomes the amino terminal residue which is then modified by covalently bound fatty acids.

\*

```

ATGGATCCAAAACTTAGCCTTTTATTAGCAGCTGGCGTACTAGCAGGTTGT
```

AGCAGCCATTCAAAATGGCGAAACCCAAATGAAATCAGACAAAATCATT  
GCTCACCGTGGTGCTAGCGTTATTACCAGAGCATACGTTAGAATCTAAAGCACTT  
GCTTTGCACAACAGGCTGATTATTAGAGCAAGATTTAGCAATGACTAAGGATGGT  
CGTTTAGTGGTTATTACGATCACTTTAGATGGCTTGACTGATGTTGCGAAAAAAA  
TTCCCACATCGTCATCGTAAGATGGCCGTTACTATGTCATCGACTTACCTTAAAA  
GAAATTCAAAGTTAGAAATGACAGAAAACTTGAAACCATGGGTGGCAGTGGTC  
AAAAGTAGTGTGGTGGATGGCCTACTGTAAGGGAAAGAATGAGACGAGCTGAGCCA  
GCAGCAGATGGGGTGGGAGCAGCATCTCGAGACCTGGAAAACATGGAGCAATCACA  
AGTAGCAATACAGCAGTACCAATGCTGCTTGCCCTGGCTGAGAACGACAAGAGGAG  
GAGGAGGTGGGTTTCCAGTCACACCTCAGGTACCTTAAGACCAATGACTTACAAG  
GCAGCTGAGATCTTAGCCACTTTTAAAGAAAAGGGGGACTGGAAAGGGGCTAATT  
CACTCCAACGAAGACAAGATATCCTGATCTGTGGATCTACCACACACAAGGCTAC  
TTCCCTGATTGGCAGAAACTACACACCAGGGCAGGGTCAGATATCCACTGACCTT  
GGATGGTGCTACAAGCTAGTACCACCAGGTTGAGCCAGATAAGGTAGAAGAGGCCAAAAA  
GGAGAGAACACCAGCTGTTACACCCTGTGAGCCTGCATGGAAATGGATGACCCTGAG  
AGAGAAGTGTAGAGTGGAGGTTGACAGCCCTAGCATTCATACGTGGCCACCATCACCATCACCAT  
TAA

Protein sequence of the processed lipidated ProtD-Nef-HIS protein (Seq. ID. No. 15)

(Amino-acids corresponding to Prot D fusion partner are in bold)

```

CSSHSSNMANTQMKSDKIIIAHRGASGYLPEHTLESKALAFAQQADYLEQDLAMTKD  

GRLVVIHDHFLDGLTDVAKKFPHRHKRDGRYYVIDFTLKEIQSLEMTENFETMGGKW  

SKSSVVGPTVRERMRAEPAADGVGAASRDLEKHGAITSSNTAATNAACAWLEAQE  

EEEVGFPVTPQVPLRPMTYKAAVDLSHFLKEGGLEGLIHSQRQDILDLWIYHTQG  

YFPDWQNYTPGPVRYPLTFGWCYKLVPVEPDKVEEANKGENTSLLHPVSLHGMDP  

EREVLEWRFDSRLAFHHVARELHPEYFKNCTSGHHHHHH.

```

⇒ LipoD-Nef-Tat-HISDNA sequence (Seq. ID. No. 16)

Nucleotides corresponding to the Prot D Fusion Partner are in bold.  
 The Lipidation Signal Sequence is underlined. After processing, the cysteine coded by the TGT codon, indicated with a star, becomes the amino terminal residue which is then modified by covalently bound fatty acids.

\*

```

ATGGATCCAAAACTTAGCCTTTCTTTATTAGCAGCTGGCGTACTAGCAGGTTGT
AGCAGCCATTCATCAAATGGGAATACCCAAATGAAATCAGACAAATCATTATT
GCTCACCGTGGGTGCTAGCGGTTATTTACCAGAGCATACGTTAGAATCTAAAGCACTT
GCGTTTGCACAACAGGCTGATTTTAGAGCAAGATTTAGCAATGACTTAAGGATGGT
CGTTAGGGTTATTCACGATCACTTTTAGATGGCTTGACTGATGTTTGCGAAAAAA
TTCCCACATCGTCATCGTAAAGATGGCCGTTACTATGTCATCGACTTTACCTAAAA
GAAATTCAAAGTTAGAAATGACAGAAACTTTGAAACCATGGGTGGCAGTGGTCA
AAAAGTAGTGTGGTGGATGGCCTACTGTTAAGGAAAGAATGAGACGAGCTGAGCC
GCAGCAGATGGGGTGGGGAGCAGCATCTCGAGACCTGGAAAACATGGGAGCATCACA
AGTAGCAATACAGCAGCTACCAATGCTGCTGTGCTGGCTAGAAGCACAAAGGGAG
GAGGAGGTGGGTTCCAGTCACACCTCAGGTACTTAAAGACCAATGACTTACAAG
GCAGCTGTAGATCTTAGCCACTTTAAAGAAAAGGGGGACTGGAAAGGGCTTAATT
CACTCCCAACGAACGAAGAATTCCTTGATCTGTGATCTACACACACAAGGCTAC
TTCCCTGATTGGCAGAACTACACACCAGGGGCCAGGGGTCAGATATCCACTGACCTTT
GGATGGGCTACAAGCTAGTACCGTTGAGCCAGATAAGGTAGAAGAGGCCAAAAA
GGAGAGAACACCAGCTGTGTTACACCGTGAGCCTGCATGGATGGATGACCCCTGAG
AGAGAAGTGTTAGAGTGGAGGTTGACAGCCGCTAGCATTCATCACGTGGCCGA
GAGCTGCATCCGGAGTACTTCAGAACTGCACTAGTGAGCCAGTAGATCCTAGACT
GAGCCCTGGAAAGCATCCAGGAAGTCAGGCCTAAAACTGCTTGTACCAATTGCTATTG
AAAAAGTGTGCTTCATTGCCAAAGTTGTTCATAACAAAGCCTTAGGCATCTCC
TATGGCAGGAAGAAGCGGAGACAGCGACGAAGACCTCTCAAGGCAGTCAGACTCAT
CAAGTTCTCTATCAAAGCAACCCACCTCCAATCCGAGGGACCCGACAGGCCCG
AAGGAAACTAGTGGCCACCCATCACCCACATCACCCATTAA
```

Protein sequence of the processed lipidated ProtD-NEF-TAT-HIS protein (Seq. ID. No. 17)

(Amino-acids corresponding to Prot D fusion partner are in bold)

```

CSSHSSNMANTQMKSDIIIIHRGASGYLPEHTLESKALAFAQQADYLEQDLAMTKD
GRLVVIHDHFLGLDGLTDVAKKFPHRHRKDGYYVIDFTLKEIQSLEMTENFETMGGKW
SKSSVVGWPVRERMRRAEPAADGVGAASRDLEKGAITSNTAATNAACAWLEAQE
EEEVGFPVTPQVPLRPMTYKAAVDLSHFLKEKGGLEGLIHSQRQDILDLWIYHTQG
YFPDWQNYTPGPGVRYPLTFGCYKLVPVEPDKEAKGENTSLLHPVSLHGMDDP
EREVLEWRFDSRLAFHHVARELHPEYFKNCTSEPVDPRLEPWKHPGSQPKTACTNCY
CKKCCFHQVCFITKALGISYGRKKRQRRRPPQGSQTHQVSLSKQPTSQRGDPTG
PKETSGHHHHHH.

```

⇒ ProtD-Nef-HIS

DNA sequence (Seq. ID. No. 18)

Nucleotides corresponding to the Prot D Fusion Partner are in bold.

```

ATGGATCCAAGCAGCCATTCATCAAATATGGCGAATACCCAAATGAAATCAGACAAA
ATCATTATTGCTCACCGTGGTGCAGCGTTATTCACAGAGCATACTGTTAGAATCT
AAAGCACTTGCGTTGCACAAACAGGCTGATTATTTAGAGCAAGATTTAGCAATGACT
AAGGATGGTCGTTAGTGGTTATTACGATCACTTTAGATGGCTTGACTGATGTT
GCGAAAAAAATTCCCACATCGTCATCGTAAAGATGGCCGTTACTATGTCATCGACTTT
ACCTTAAAAGAAATTCAAAGTTAGAAATGACAGAAAACTTGAAACCATTGGTGGC
AAGTGGTCAAAAGTAGTGTGGTTGGATGGCCTACTGTAAGGGAAAGAATGAGACGA
GCTGAGGCCAGCAGCAGATGGGTGGGAGCAGCATCTCGAGACCTGGAAAAACATGGA
GCAATCACAAAGTAGCAATACAGCAGCTACCAATGCTGCTTGCGCTGGCTAGAAC
CAAGAGGAGGAGGAGGTGGTTTCAGTCACACCTCAGGTACCTTAAGACCAATG
ACTTACAAGGAGCTGTAGATCTTAGCCACTTTTAAAGAAAAGGGGGACTGGAA
GGGCTAATTCACTCCCAACGAAGACAAGATATCCTGATCTGTGGATCTACCACACA
CAAGGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGTCAGATATCCA
CTGACCTTGGATGGTGCCTACAAGCTAGTACCAAGTTGAGGCCAGATAAGGTAGAAGAG
GCCAATAAAAGGAGAGAACACCAGCTGTTACACCCCTGTGAGCCTGCATGGAATGGAT
GACCCTGAGAGAGAAGTGTAGAGTGGAGGTTGACAGCCGCCTAGCATTTCATCAC
GTGGCCCGAGAGCTGCATCCGGAGTACTTCAAGAACTGCACTAGTGGCCACCATCAC
CATCACCATTA

```

Protein sequence (Seq. ID. No. 19)

(Amino-acids corresponding to Prot D fusion partner are in bold)

```

MDPSSHSSNMANTQMKSDKIIIAHRGASGYLPEHTLESKALAFQQADYL
EQDLAMTKDGRLVVIHDHFLDGLTDVAKKFPHRHRKDGRYYVIDFTLK
EIQSLEMTENFETMGGKWSKSSVVGWPTVRERMRAEPAADGVGAASRDL
EKHGAITSSNTAATNAACAWLEAQEEEVGFVTPQVPLRPMTYKAAVDLSH
FLKEKGGLEGLIHSQRRQDILDWIYHTQGYFPDWQNYTPGPGVRYPLTFGW
CYKLVPVEPDKVEEANKGENTSLLHPVSLHGMDDPEREVLEWRFDSRLAFH
HVARELHPEYFKNCTSGHHHHHHH.

```

⇒ ProtD-Nef-Tat-HIS

DNA sequence (Seq. ID. No. 20)

7/17

Nucleotides corresponding to the Prot D Fusion Partner are in bold.

ATGGATCCAAGCAGCCATTCATCAAATATGGCGAATA**CCAAATGAAATCAGACAAA**  
 ATCATTATTGCTCACCGTGGTGTAGCGGTTATTTACCA**GAGCATA**CGTTAGAATCT  
 AAAGCACTTGCGTTTGACA**ACACAGGCTGATT**ATTTAGAGCAAGATTAGCAATGACT  
 AAGGATGGTCGTTAGTGGTTATT**CACGATCACTTTAGATGGCTTGACTGATGTT**  
 GCGAAAAAATTCCCACATCGTCATCGTAAAGATGGCCGTTACTATGT**CATCGACTTT**  
 ACCTTAAAAGAAATT**CAAAGTTAGAAATGACAGAAA**ACTTTGAAACC**ATGGTGGC**  
 AAGTGGTCAAAAAGTAGTGTGGTTGGATGGCCTACTGT**AAGGGAAAGAATGAGACGA**  
 GCTGAGCCAGCAGCAGATGGGGTGGGAGCAGC**ATCTCGAGACCTGGAAAAACATGGA**  
 GCAATCACAAGTAGCAATACAGCAGCTACCAATGCTGCTTGTGCCTGGCTAGAAGCA  
 CAAGAGGAGGAGGAGGTGGGTTTCCAGTCACAC**CTCAGGTACCTTTAAGACCAATG**  
 ACTTACAAGGCAGCTGTAGATCTT**AGCCACTTTAAAAGAAAAGGGGGACTGGAA**  
 GGGCTAAITC**ACTCCAACGAAGACAAGAT**TCTGATCTGTGGATCTACCACACA  
 CAAGGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGGT**CAGATATCCA**  
 CTGACCTTGGATGGTGT**ACAAGCTAGTACCA**GGCAGATAAGGTAGAAGAG  
 GCCAATAAAGGAGAGAACACCAGCTTGT**TACACCTGTGAGCCTGCATGGAATGGAT**  
 GACCTGAGAGAGAAGTGTAGAGTGGAGGTTGACAGCCGCTAGCATT**CATCAC**  
 GTGGCCCGAGAGCTGC**ATCCGGAGTACTTCAAGAACTGC**ACTAGTGAGCCAGTAGAT  
 CCTAGACTAGAGCCCTGGAAAG**CATCCAGGAAGTCAGCCTAAACTGCTTG**TACCAAT  
 TGCTATTGTA**AAAAGTGTGCTTCATTGCAAGTTGTT**CATAACAAA**AGCCTTA**  
 GGCATCTCTATGGCAGGAAGAAGCGGAGACAGC**GACGAAGACCTCCTCAAGGCAGT**  
 CAGACTCATCAAGTTCTATCAA**AGCAACCCACCTCCAATCCGAGGGGACCG**  
 ACAGGCCGAAGGAA**ACTAGTGGCCACC**ATCACC**ATCACCATTAA**

Protein sequence (Seq. ID. No. 21)

(Amino-acids corresponding to Prot D fusion partner are in bold)

MDPSSHSNNMANTQM**KSDKIIIAHRGASGYLPEHTLES**KALAF**AQQADYLEQDLAMT**  
 KDGR**LVVIHDHFLDGLTDVAKKFPHRHRKDGRYYVIDFTLKEIQSLEM**TENFETMGG  
 KWSKSSVVGWPTVRERMRRAEP**AADGVGAASRDLEKHGAITSSNTAATNAACAWLEA**  
 QEEEVGF**PVTPQVPLRPMTYKAAVDLSHFLKEKGGLIHSQRQDILDLWIYHT**  
 QGYFPDWQNYTPGPVGVR**YPLTFGWCYKLVPVEPDVKVEEANKGENTSLLHPVSLHGMD**  
 DPEREV**LEWRFD**SLAFHVAREL**HPEYFKNCTSEPVDPRLEPWKHPGSQPKTACTN**  
 CYCKKCCFH**CQVCFITKALGI**SYGRKKRRQRRPPQGSQTHQV**SLSKQPTSQRGDPT**  
 TGPKETSGHHHHHH.

**⇒ Tat-MUTANT-HIS**

DNA sequence (Seq. ID. No. 22)

ATGGAGCCAGTAGATCCTAGACTAGAGCCCTGGAAGCATC	40
CAGGAAGTCAGCCTAAAACGTGCTTGTACCAATTGCTATTG	80
TAAAAAGTGTGCTTCATTGCCAAGTTGTTCTATAACA	120
GCTGCCTTAGGCATCTCTATGGCAGGAAGAACGGAGAC	160
AGCGACGAAGACCTCTCAAGGCAGTCAGACTCATCAAGT	200
TTCTCTATCAAAGCAACCCACCTCCAATCAAAGGGAG	240
CCGACAGGCCGAAGGAAACTAGTGGCCACCATCACCATC	280
ACCATTAA	288

Protein sequence(Seq. ID. No. 23)

Mutated amino-acids in Tat sequences are in bold.

MEPVDPRLPWKHPGSQPKTACTNCYCKKCCFHQCQVCFIT	40
<b>AALGISYGRKKRQRQQPPQGSQTHQVSLSKQPTSQSKE</b>	80
PTGPKETSGHHHHHH.	95

⇒*Nef-Tat-Mutant-HIS*DNA sequence(Seq. ID. No. 24)

ATGGGTGGCAAGTGGTCAAAAAGTAGTGTGGTTGGATGGC	40
CTACTGTAAGGAAAGAATGAGACGAGCTGAGCCAGCAGC	80
AGATGGGGTGGGAGCAGCATCTCGAGACCTGGAAAACAT	120
GGAGCAATCACAAAGTAGCAATAACAGCAGCTACCAATGCTG	160
CTTGTGCCTGGCTAGAAGCACAAGAGGAGGAGGAGGTGGG	200
TTTCCAGTCACACCTCAGGTACCTTAAGACCAATGACT	240
TACAAGGCAGCTGTAGATCTTAGCCACTTTAAAGAAA	280
AGGGGGGACTGGAAGGGCTAATTCACTCCAACGAAAGACA	320
AGATATCCTTGATCTGGATCTACCACACACAAGGCTAC	360
TTCCCTGATTGGCAGAACTACACACCAGGGCAGGGTCA	400
GATATCCACTGAÇCTTGGATGGCTACAAGCTAGTACC	440
AGTTGAGCCAGATAAGGTAGAAGAGGCCATAAAGGAGAG	480
AACACCAGCTTGTACACCCTGTGAGCCTGCATGGAATGG	520
ATGACCCCTGAGAGAGAAGTGTAGAGTGGAGGTTGACAG	560
CCGCCTAGCATTTCATCACGTGGCCCGAGAGCTGCATCCG	600
GAGTACTTCAAGAACTGCACTAGTGAGCCAGTAGATCCTA	640
GAATAGAGCCCTGGAAGCATCCAGGAAGTCAGCCTAAAAC	680
TGCTTGTACCAATTGCTATTGTAAGGTGTGCTTTCAT	720
TGCCAAGTTGTTCTATAACAGCTGCCTTAGGCATCTCCT	760
ATGGCAGGAAGAACGGAGACAGCGACGAAGACCTCCTCA	800
AGGCAGTCAGACTCATCAAGTTCTATCAAAGCAACCC	840
ACCTCCCAATCAAAGGGAGCCGACAGGCCGAAGGAAA	880
CTAGTGGCCACCATCACCACCATCACCATTAA	909

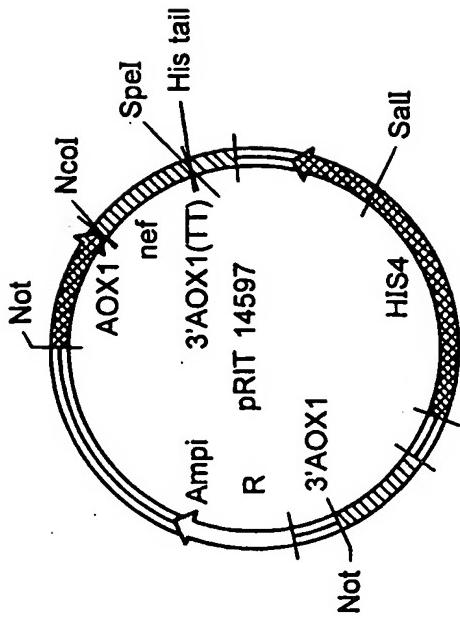
9/17

Protein sequence (Seq. ID. No. 25)

Mutated amino-acids in Tat sequence are in bold.

MGGKWSKSSVVGWPTVRERMRAEPAADGVGAASRDLEKH 40  
GAITSSNTAATNAACAWLEAQEEEEVGFPVTPQVPLRPMT 80  
YKAAVDSLHFLKEKGGLIHSQRQDILDWIYHTQGY 120  
FPDWQNYTPGPGVRYPLTFGWCYKLPVEPDKVEEANKGE 160  
NTSLLHPVSLHGMDPEREVLEWRFDSRLAFHHVARELHP 200  
EYFKNCTSEPVDPRLEPWKHPGSQPKTACTNCYCKCCFH 240  
CQVCFITAALGISYGRKKRQRRPPQGSQTHQVSLSKQP 280  
TSQSKGEPTGPKETSGHHHHHH . 302

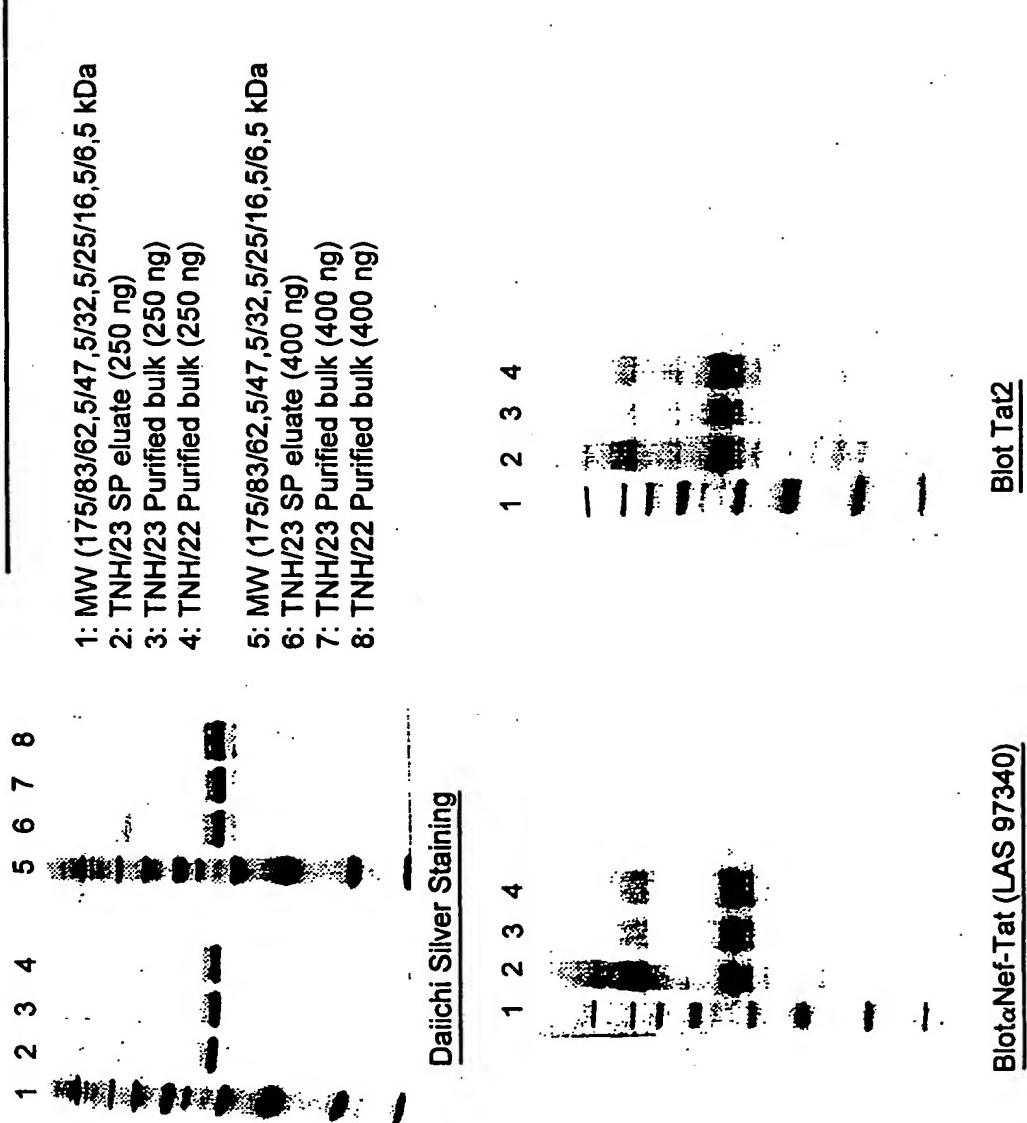
**Fig . 3** Map of pRIT14597 integrative vector



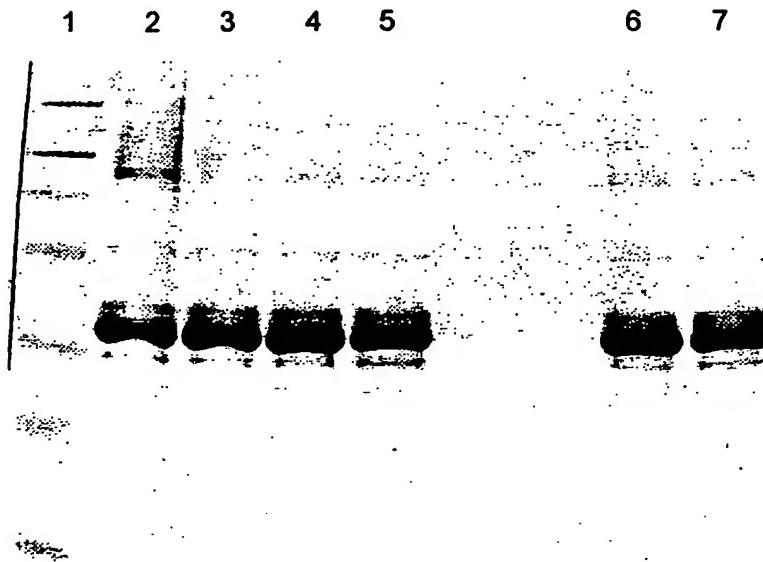
MCS POLYLINKER: *nef* gene inserted between NcoI and SpeI sites.

<i>Acu II</i>	<i>Nco I</i>	<i>Spe I</i>	<i>Eco RI</i>
TTCGAA ACC.ATGGCCGGACTAGT	GGC.CAC.CAT.CAC.CAT.TAA.CGGAAATTC		
		Thr . Ser . Gly . His . His . His . His . His	

The amino acid sequence of Figure 3 relates to Seq. ID no. 27 and the nucleic acid sequence of Figure 3 relates to Seq. ID. No.26.

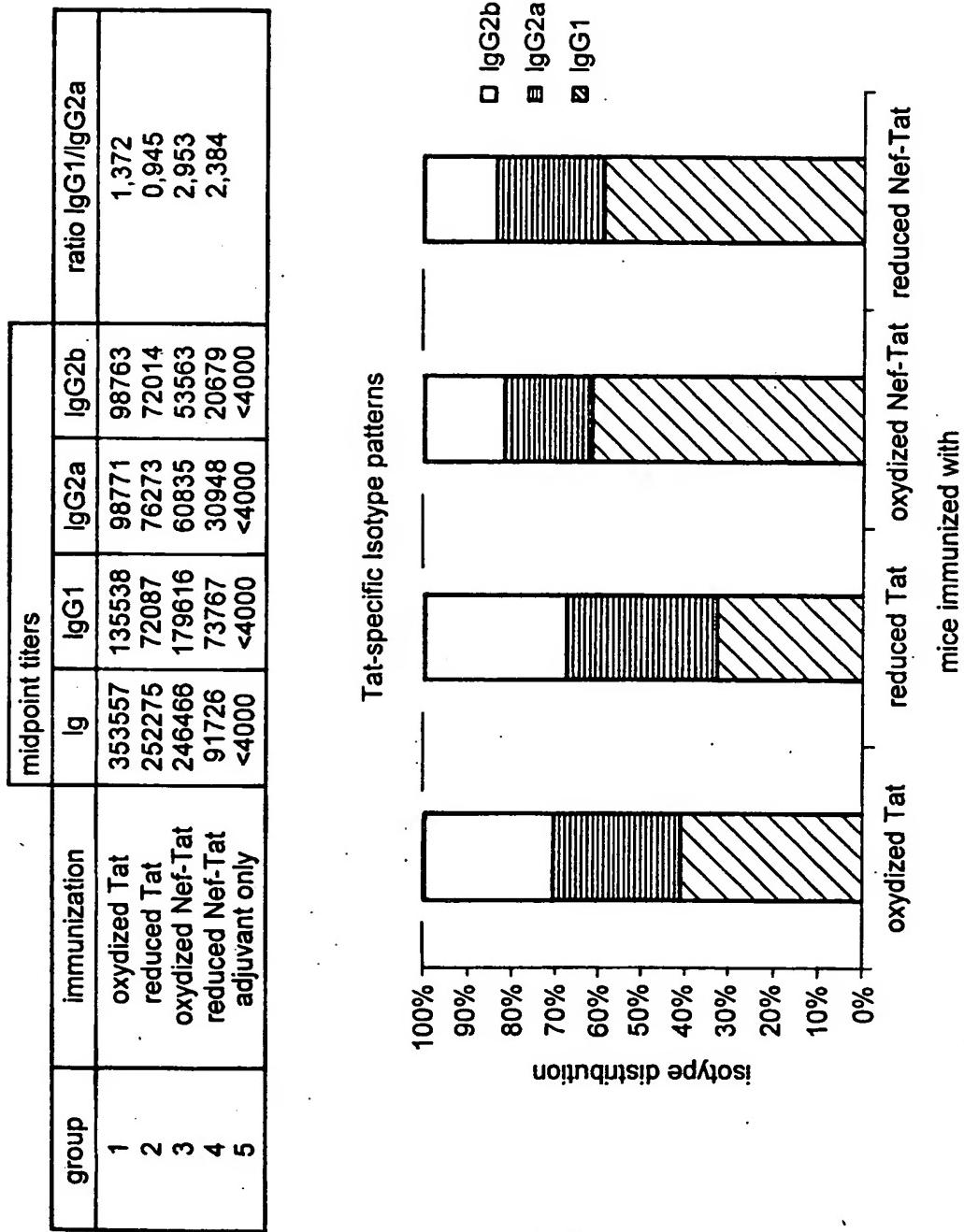
**Fig . 4 SDS-PAGE: Nef-Tat-his fusion protein**

**Fig . 5 SDS-PAGE: Nef-Tat-his fusion protein**



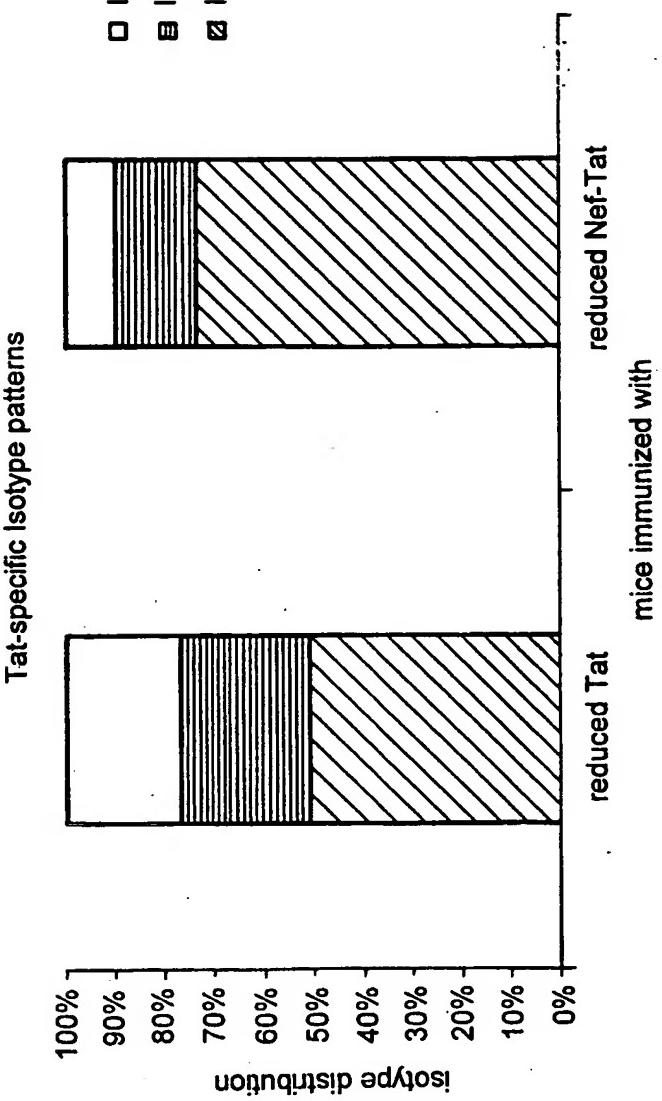
Coomassie blue G250

- 1: MW (175/83/62,5/47,5/32,5/25/16,5/6,5 kDa)
- 2: TNH/23 SP eluate (4 µg)
- 3: TNH/23 Superdex200 eluate (4 µg)
- 4: TNH/23 Purified bulk (4 µg)
- 5: TNH/22 Purified bulk (4 µg)
- 6: TNH/23 Purified bulk (4 µg) / non reducing conditions
- 7: TNH/22 Purified bulk (4 µg) / non reducing conditions

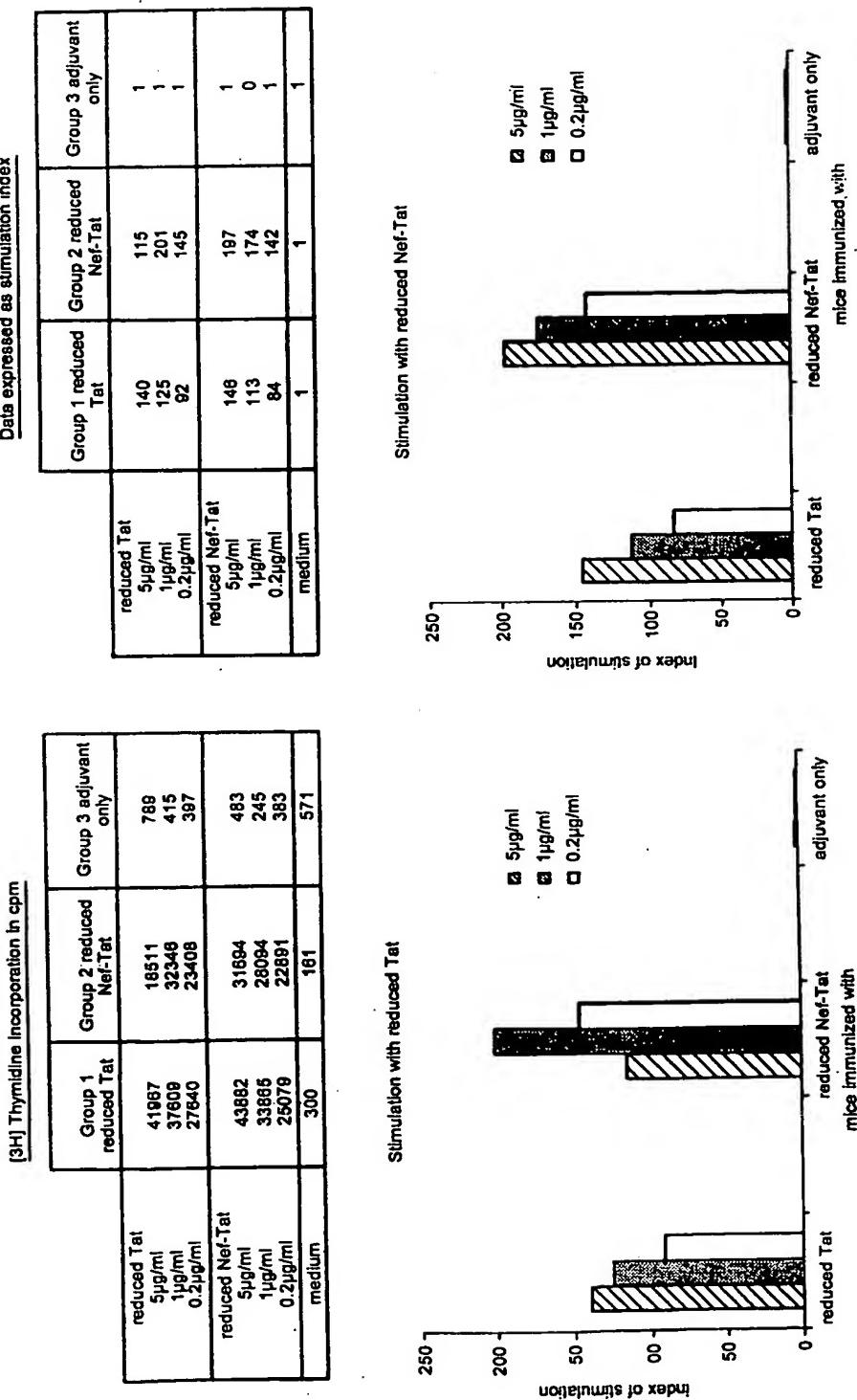
**Fig. 6A** Tat-specific antibody titers and Isotypes

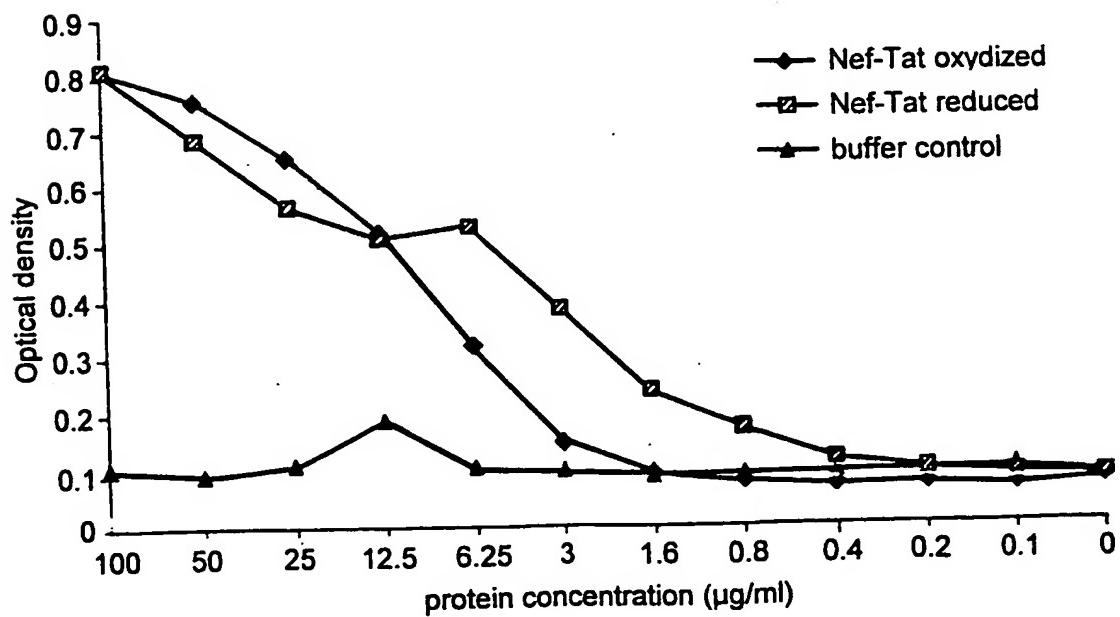
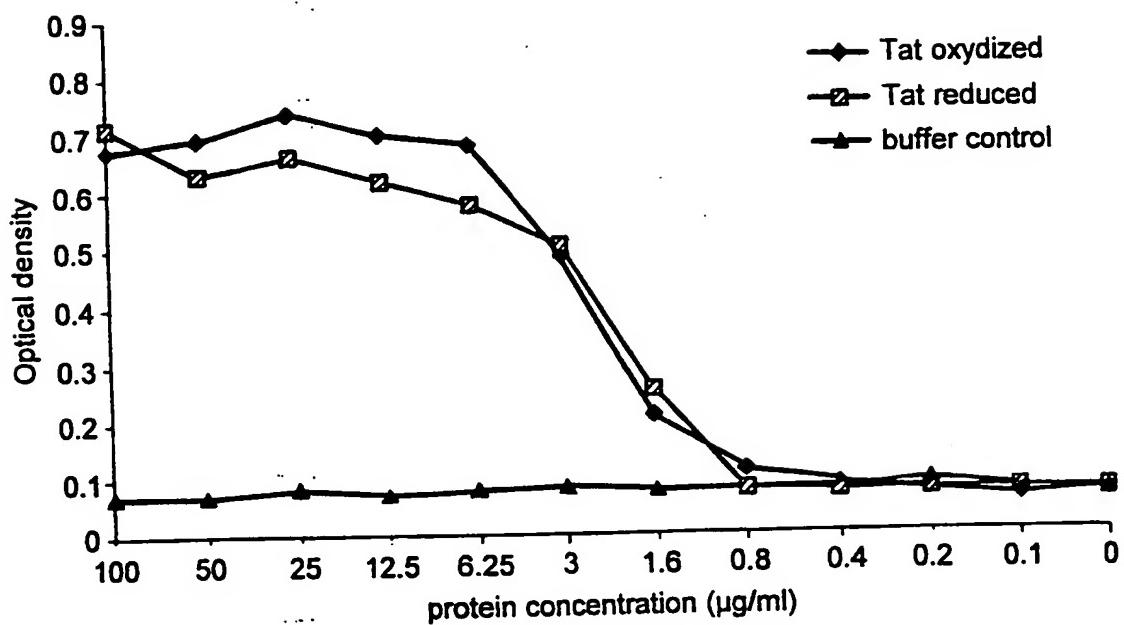
**Fig. 6B Tat-specific antibody titers and Isotypes**

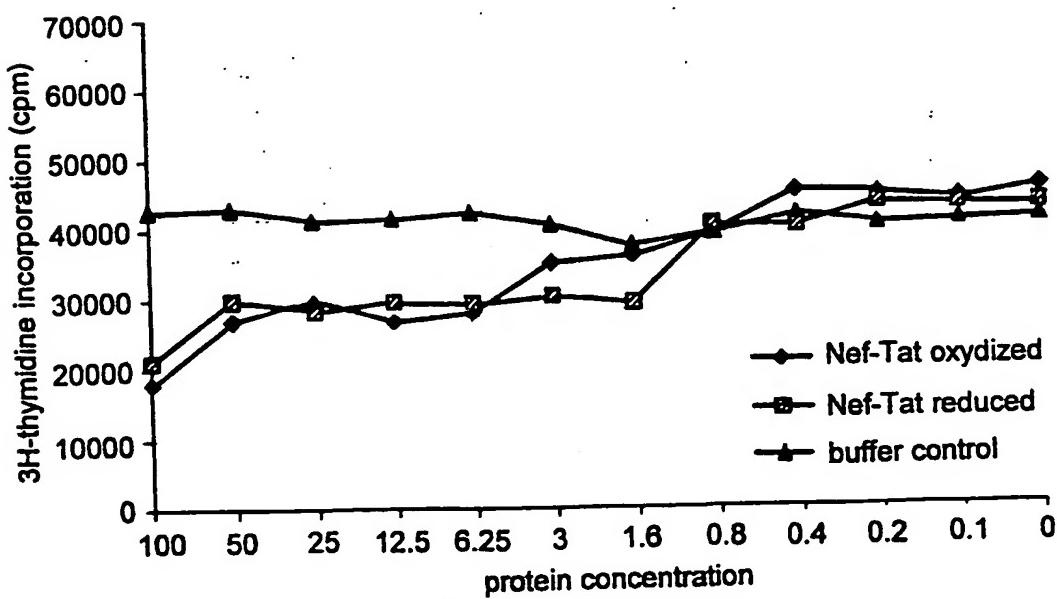
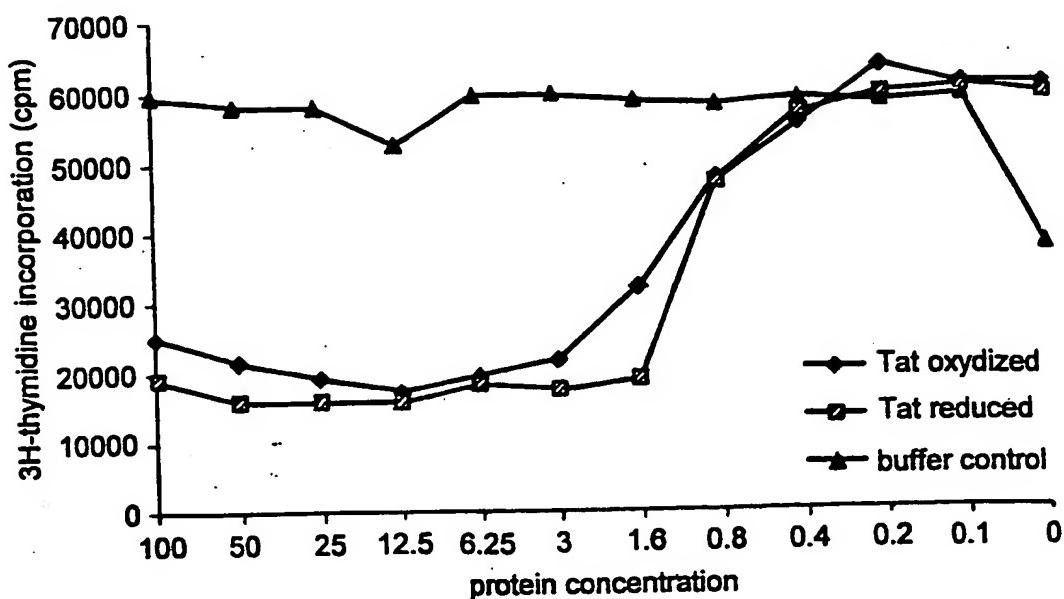
group	immunization	midpoint titers			ratio IgG1/IgG2a
		Ig	IgG1	IgG2a	
1	reduced Tat	212799	123242	62697	1.966
2	reduced Nef-Tat	75676	84046	11692	4.556
3	adjuvant only	<4000	<4000	<4000	



**Fig. 7** Antigen-specific lymphoproliferative response of pooled lymph node cells



**Fig. 8 Cell binding assay**

**Fig. 9** Inhibition of cell growth

## SEQUENCE LISTING

## (1) GENERAL INFORMATION

(i) APPLICANT: SmithKline Beecham Biologicals S.A.

(ii) TITLE OF THE INVENTION: Vaccine

(iii) NUMBER OF SEQUENCES: 27

(iv) CORRESPONDENCE ADDRESS:

- (A) ADDRESSEE: SmithKline Beecham
- (B) STREET: Two New Horizons Court
- (C) CITY: Brentford
- (D) STATE:
- (E) COUNTRY: Middx, UK
- (F) ZIP: TW8 9EP

(v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Diskette
- (B) COMPUTER: IBM Compatible
- (C) OPERATING SYSTEM: DOS
- (D) SOFTWARE: FastSEQ for Windows Version 2.0

(vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER:
- (B) FILING DATE: 26-SEP-1997
- (C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER:
- (B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: Bor, Fiona R
- (B) REGISTRATION NUMBER:
- (C) REFERENCE/DOCKET NUMBER:

(ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: 0181 975 2817
- (B) TELEFAX: 0181 975 6141
- (C) TELEX:

## (2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATCGTCCATG .GGT.GGC.A AG.TGG.T

28

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CGGCTACTAG TGCAGTTCTT GAA

23

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATCGTACTAG T.GAG.CCA. GTA.GAT.C

29

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CGGCTACTAG TTTCCTTCGG GCCT

24

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ATCGTCCATG GAGCCAGTAG ATC

23

## (2) INFORMATION FOR SEQ ID NO:6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 441 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

ATGGATCCAA	AAACTTTAGC	CCTTTCTTA	TTAGCAGCTG	GCGTACTAGC	AGGTTGTAGC	60
AGCCATTCA	CAAATATGGC	GAATAACCAA	ATGAAATCAG	ACAAAATCAT	TATTGCTCAC	120
CGTGGTGCTA	GCGGTTATTT	ACCAAGAGCAT	ACGTTAGAAT	CTAAAGCACT	TGCTTTTGCA	180
CAACACGGCTG	ATTATTTAGA	GCAAGATTAA	GCAATGACTA	AGGATGGTCG	TTTAGTGGTT	240
ATTCAACGATC	ACTTTTAA	TGGCTTGACT	GATGTTGCAG	AAAAAATTCCC	ACATCGTCAT	300
CGTAAAGATG	GCCGTTACTA	TGTCATCGAC	TTTACCTTAA	AAGAAATTCA	AGTTTAGAA	360
ATGACAGAAA	ACTTTGAAAC	CATGCCACG	TGTGATCAGA	GCTCAACTAG	TGGCCACCAT	420
CACCATCACC	ATTAATCTAG	A				441

## (2) INFORMATION FOR SEQ ID NO:7:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 144 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met	Asp	Pro	Lys	Thr	Leu	Ala	Leu	Ser	Leu	Leu	Ala	Ala	Gly	Val	Leu
1				5					10					15	
Ala	Gly	Cys	Ser	Ser	His	Ser	Ser	Asn	Met	Ala	Asn	Thr	Gln	Met	Lys
					20			25					30		
Ser	Asp	Lys	Ile	Ile	Ile	Ala	His	Arg	Gly	Ala	Ser	Gly	Tyr	Leu	Pro
					35			40				45			
Glu	His	Thr	Leu	Glu	Ser	Lys	Ala	Leu	Ala	Phe	Ala	Gln	Gln	Ala	Asp
					50			55				60			
Tyr	Leu	Glu	Gln	Asp	Leu	Ala	Met	Thr	Lys	Asp	Gly	Arg	Leu	Val	Val
					65			70			75		80		
Ile	His	Asp	His	Phe	Leu	Asp	Gly	Leu	Thr	Asp	Val	Ala	Lys	Lys	Phe
					85			90				95			
Pro	His	Arg	His	Arg	Lys	Asp	Gly	Arg	Tyr	Tyr	Val	Ile	Asp	Phe	Thr
					100			105			110				
Leu	Lys	Glu	Ile	Gln	Ser	Leu	Glu	Met	Thr	Glu	Asn	Phe	Glu	Thr	Met
					115			120			125				
Ala	Thr	Cys	Asp	Gln	Ser	Ser	Thr	Ser	Gly	His	His	His	His	His	His
					130			135			140				

## (2) INFORMATION FOR SEQ ID NO:8:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 648 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

ATGGGTGGCA	AGTGGTCAAA	AAGTAGTGTG	GTTGGATGGC	CTACTGTAAG	GGAAAGAATG	60
AGACGAGCTG	AGCCAGCAGC	AGATGGGTG	GGAGCAGCAT	CTCGAGACCT	GGAAAAACAT	120
GGAGCAATCA	CAAGTAGCAA	TACAGCAGCT	ACCAATGCTG	CTTGTGCCCTG	GCTAGAAGCA	180
CAAGAGGAGG	AGGAGGTGGG	TTTCCAGTC	ACACCTCAGG	TACCTTTAAG	ACCAATGACT	240
TACAAGGCAG	CTGTAGATCT	TAGCCACTT	TTAAAAGAAA	AGGGGGGACT	GGAAGGGCTA	300
ATTCACTCCC	AACGAAGACA	AGATATCCTT	GATCTGTGGA	TCTACCACAC	ACAAGGCTAC	360
TTCCCTGATT	GGCAGAACTA	CACACCAGGG	CCAGGGGTCA	GATATCCACT	GACCTTTGGA	420
TGGTGTACAA	AGCTAGTACC	AGTGAGCCA	GATAAGGTAG	AAGAGGCCAA	TAAAGGAGAG	480
AACACCAGCT	TGTTACACCC	TGTGAGCCTG	CATGGAATGG	ATGACCCCTGA	GAGAGAAGTG	540
TTAGAGTGGA	GGTTTGACAG	CCGCTAGCA	TTTCATCACG	TGGCCCGAGA	GCTGCATCCG	600
GAGTACTTCA	AGAACTGCAC	TAGTGGCCAC	CATCACCATC	ACCATTAA		648

## (2) INFORMATION FOR SEQ ID NO:9:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 216 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Met	Gly	Gly	Lys	Trp	Ser	Lys	Ser	Ser	Val	Val	Gly	Trp	Pro	Thr	Val
1					5				10				15		
Arg	Glu	Arg	Met	Arg	Arg	Ala	Glu	Pro	Ala	Ala	Asp	Gly	Val	Gly	Ala
						20			25				30		
Ala	Ser	Arg	Asp	Leu	Glu	Lys	His	Gly	Ala	Ile	Thr	Ser	Ser	Asn	Thr
						35			40			45			
Ala	Ala	Thr	Asn	Ala	Ala	Cys	Ala	Trp	Leu	Glu	Ala	Gln	Glu	Glu	Glu
						50			55			60			
Glu	Val	Gly	Phe	Pro	Val	Thr	Pro	Gln	Val	Pro	Leu	Arg	Pro	Met	Thr
						65			70			75		80	
Tyr	Lys	Ala	Ala	Val	Asp	Leu	Ser	His	Phe	Leu	Lys	Glu	Lys	Gly	Gly
						85			90			95			
Leu	Glu	Gly	Leu	Ile	His	Ser	Gln	Arg	Arg	Gln	Asp	Ile	Leu	Asp	Leu
						100			105			110			
Trp	Ile	Tyr	His	Thr	Gln	Gly	Tyr	Phe	Pro	Asp	Trp	Gln	Asn	Tyr	Thr
						115			120			125			
Pro	Gly	Pro	Gly	Val	Arg	Tyr	Pro	Leu	Thr	Phe	Gly	Trp	Cys	Tyr	Lys
						130			135			140			
Leu	Val	Pro	Val	Glu	Pro	Asp	Lys	Val	Glu	Glu	Ala	Asn	Lys	Gly	Glu
						145			150			155		160	
Asn	Thr	Ser	Leu	Leu	His	Pro	Val	Ser	Leu	His	Gly	Met	Asp	Asp	Pro
						165			170			175			
Glu	Arg	Glu	Val	Leu	Glu	Trp	Arg	Phe	Asp	Ser	Arg	Leu	Ala	Phe	His
						180			185			190			
His	Val	Ala	Arg	Glu	Leu	His	Pro	Glu	Tyr	Phe	Lys	Asn	Cys	Thr	Ser
						195			200			205			
Gly	His	His	His	His	His										
						210			215						

## (2) INFORMATION FOR SEQ ID NO:10:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 288 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

ATGGGAGCCAG TAGATCCTAG ACTAGAGCCC	TGGAAGCATC CAGGAAGTCA GCCTAAAAC	60
GCTGTACCA ATTGCTATTG TAAAAAGTGT	TGCTTTCAATT GCCAAGTTTG TTTCATAACA	120
AAAGCCTTAG GCATCTCTA TGGCAGGAAG AAGCGGAGAC	AGCGACGAAG ACCTCCTCAA	180
GGCAGTCAGA CTCATCAAGT TTCTCTATCA AAGCAACCCA	CCTCCCAATC CCGAGGGGAC	240
CCGACAGGCC CGAAGGAAAC TAGTGGCAC	CATCACCATC ACCATTAA	288

## (2) INFORMATION FOR SEQ ID NO:11:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 96 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Met Glu Pro Val Asp Pro Arg Leu Glu Pro Trp Lys His Pro Gly Ser			
1	5	10	15
Gln Pro Lys Thr Ala Cys Thr Asn Cys Tyr Cys Lys Lys Cys Cys Phe			
20	25	30	
His Cys Gln Val Cys Phe Ile Thr Lys Ala Leu Gly Ile Ser Tyr Gly			
35	40	45	
Arg Lys Lys Arg Arg Gln Arg Arg Arg Pro Pro Gln Gly Ser Gln Thr			
50	55	60	
His Gln Val Ser Leu Ser Lys Gln Pro Thr Ser Gln Ser Arg Gly Asp			
65	70	75	80
Pro Thr Gly Pro Lys Glu Thr Ser Gly His His His His His His			
85	90	95	

## (2) INFORMATION FOR SEQ ID NO:12:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 909 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

ATGGGTGGCA AGTGGTCAAA AAGTAGTGTG GTTGGATGGC CTACTGTAAG GGAAAGAAC	60
AGACGAGCTG AGCCAGCAGC AGATGGGTG GGAGCAGCAT CTCGAGACCT GGAAAAACAT	120
GGAGCAATCA CAAGTAGCAA TACAGCAGCT ACCAATGCTG CTTGTGCCTG GCTAGAAGCA	180
CAAGAGGAGG AGGAGGTGGG TTTCCAGTC ACACCTCAGG TACCTTTAAG ACCAATGACT	240
TACAAGGCAG CTGTAGATCT TAGCCACTTT TTAAAAGAAA AGGGGGGACT GGAAGGGCTA	300
ATTCACTCCC AACGAAGACA AGATATCCTT GATCTGTGGA TCTACCACAC ACAAGGCTAC	360

TTCCCTGATT	GGCAGAACTA	CACACCAGGG	CCAGGGTCA	GATATCCACT	GACCTTTGGA	420
TGGTGCTACA	AGCTAGTACC	AGTTGAGCCA	GATAAGGTAG	AAGAGGCCAA	TAAAGGAGAG	480
AACACCAAGCT	TGTTACACCC	TGTGAGCTG	CATGGAATGG	ATGACCCCTGA	GAGAGAAGTG	540
TTAGAGTGGG	GGTTTGACAG	CCGCCTAGCA	TTTCATCACG	TGGCCCGAGA	GCTGCATCCG	600
GAGTACTTCA	AGAACTGCAC	TAGTGAGCCA	GTAGATCCTA	GACTAGAGCC	CTGGAAGCAT	660
CCAGGAAGTC	AGCCTAAAC	TGCTTGTACC	AATTGCTATT	GTAAAAGTG	TTGCTTTCAT	720
TGCCAAGTTT	GTTTCATAAAC	AAAAGCCTTA	GGCATCTCT	ATGGCAGGAA	GAAGCGGAGA	780
CAGCGACGAA	GACCTCCTCA	AGGCAGTCAG	ACTCATCAAG	TTTCTCTATC	AAAGCAACCC	840
ACCTCCCAAT	CCCGAGGGGA	CCCGACAGGC	CCGAAGGAAA	CTAGTGGCCA	CCATCACCAT	900
CACCATTA						909

## (2) INFORMATION FOR SEQ ID NO:13:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 303 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Met	Gly	Gly	Lys	Trp	Ser	Lys	Ser	Ser	Val	Val	Gly	Trp	Pro	Thr	Val
1					5				10				15		
Arg	Glu	Arg	Met	Arg	Arg	Ala	Glu	Pro	Ala	Ala	Asp	Gly	Val	Gly	Ala
						20			25				30		
Ala	Ser	Arg	Asp	Leu	Glu	Lys	His	Gly	Ala	Ile	Thr	Ser	Ser	Asn	Thr
						35			40				45		
Ala	Ala	Thr	Asn	Ala	Ala	Cys	Ala	Trp	Leu	Glu	Ala	Gln	Glu	Glu	
						50			55			60			
Glu	Val	Gly	Phe	Pro	Val	Thr	Pro	Gln	Val	Pro	Leu	Arg	Pro	Met	Thr
						65			70			75			80
Tyr	Lys	Ala	Ala	Val	Asp	Leu	Ser	His	Phe	Leu	Lys	Glu	Lys	Gly	
						85			90			95			
Leu	Glu	Gly	Leu	Ile	His	Ser	Gln	Arg	Arg	Gln	Asp	Ile	Leu	Asp	Leu
						100			105			110			
Trp	Ile	Tyr	His	Thr	Gln	Gly	Tyr	Phe	Pro	Asp	Trp	Gln	Asn	Tyr	Thr
						115			120			125			
Pro	Gly	Pro	Gly	Val	Arg	Tyr	Pro	Leu	Thr	Phe	Gly	Trp	Cys	Tyr	Lys
						130			135			140			
Leu	Val	Pro	Val	Glu	Pro	Asp	Lys	Val	Glu	Glu	Ala	Asn	Lys	Gly	
						145			150			155			160
Asn	Thr	Ser	Leu	Leu	His	Pro	Val	Ser	Leu	His	Gly	Met	Asp	Asp	Pro
						165			170			175			
Glu	Arg	Glu	Val	Leu	Glu	Trp	Arg	Phe	Asp	Ser	Arg	Leu	Ala	Phe	His
						180			185			190			
His	Val	Ala	Arg	Glu	Leu	His	Pro	Glu	Tyr	Phe	Lys	Asn	Cys	Thr	Ser
						195			200			205			
Glu	Pro	Val	Asp	Pro	Arg	Leu	Glu	Pro	Trp	Lys	His	Pro	Gly	Ser	Gln
						210			215			220			
Pro	Lys	Thr	Ala	Cys	Thr	Asn	Cys	Tyr	Cys	Lys	Lys	Cys	Cys	Phe	His
						225			230			235			240
Cys	Gln	Val	Cys	Phe	Ile	Thr	Lys	Ala	Leu	Gly	Ile	Ser	Tyr	Gly	Arg
						245			250			255			
Lys	Lys	Arg	Arg	Gln	Arg	Arg	Arg	Pro	Pro	Gln	Gly	Ser	Gln	Thr	His
						260			265			270			
Gln	Val	Ser	Leu	Ser	Lys	Gln	Pro	Thr	Ser	Gln	Ser	Arg	Gly	Asp	Pro

275	280	285
Thr Gly Pro Lys Glu Thr Ser Gly His His His His His His His		
290	295	300

## (2) INFORMATION FOR SEQ ID NO:14:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1029 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

ATGGATCAA AAACTTAGC CCTTTCTTA TTAGCAGCTG GCGTACTAGC AGGTTGTAGC	60
AGCCATTCA CAAATATGGC GAATACCCAA ATGAAATCAT ACAAAATCAT TATTGCTCAC	120
CGTGGTGCTA CGGGTTATT ACCAGAGCAT ACGTTAGAAT CTAAAGCACT TGCTTTGCA	180
CAACAGGCTG ATTATTTAGA GCAAGATTG GCAATGACTA AGGATGGTCG TTAGTGGTT	240
ATTCAAGATC ACTTTTAAAGA TGGCTTGACT GATGTTGCAG AAAAATTCCC ACATCGTCAT	300
CGTAAAGATG GCGTTACTA TGTCATCGAC TTTACCTAA AAGAAATTCA AAGTTTAGAA	360
ATGACAGAAA ACTTTGAAAC CATGGGTGGC AAGTGGTCAA AAAGTAGTGT GGTTGGATGG	420
CCTACTGTAA GGGAAAGAAT GAGACGAGCT GAGCCAGCAG CAGATGGGGT GGGAGCAGCA	480
TCTCGAGACC TGGAAAAACA TGGAGCAATC ACAAGTAGCA ATACAGCAGC TACCAATGCT	540
GCTTGTGCCT GGCTAGAACG ACAAGAGGAG GAGGAGGTGG GTTTCCAGT CACACCTCAG	600
GTACCTTAA GACCAATGAC TTACAAGGC GCTGTAGATC TTAGCCACTT TTTAAAAGAA	660
AAGGGGGGAC TGGAAAGGCT AATTCACTCC CAACGAAGAC AAGATATCCT TGATCTGTGG	720
ATCTACCACA CACAAGGCTA CTTCCCTGAT TGGCAGAACT ACACACCAGG GCCAGGGGTC	780
AGATATCCAC TGACCTTGG ATGGTGTAC AAGCTAGTAC CAGTTGAGCC AGATAAGGTA	840
GAAGAGGCCA ATAAAGGAGA AACACCAGC TTGTTACACC CTGTGAGCCT GCATGGAATG	900
GATGACCCCTG AGAGAGAAGT GTTAGAGTGG AGGTTTGACA GCCGCCTAGC ATTTCATCAC	960
GTGGCCCGAG AGCTGCATCC GGAGTACTTC AAGAACTGCA CTAGTGGCCA CCATCACCAT	1020
CACCATTA	1029

## (2) INFORMATION FOR SEQ ID NO:15:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 325 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Cys Ser Ser His Ser Ser Asn Met Ala Asn Thr Gln Met Lys Ser Asp	
1 5 10 15	
Lys Ile Ile Ile Ala His Arg Gly Ala Ser Gly Tyr Leu Pro Glu His	
20 25 30	
Thr Leu Glu Ser Lys Ala Leu Ala Phe Ala Gln Gln Ala Asp Tyr Leu	
35 40 45	
Glu Gln Asp Leu Ala Met Thr Lys Asp Gly Arg Leu Val Val Ile His	
50 55 60	
Asp His Phe Leu Asp Gly Leu Thr Asp Val Ala Lys Lys Phe Pro His	
65 70 75 80	
Arg His Arg Lys Asp Gly Arg Tyr Tyr Val Ile Asp Phe Thr Leu Lys	
85 90 95	

Glu Ile Gln Ser Leu Glu Met Thr Glu Asn Phe Glu Thr Met Gly Gly  
 100 105 110  
 Lys Trp Ser Lys Ser Ser Val Val Gly Trp Pro Thr Val Arg Glu Arg  
 115 120 125  
 Met Arg Arg Ala Glu Pro Ala Ala Asp Gly Val Gly Ala Ala Ser Arg  
 130 135 140  
 Asp Leu Glu Lys His Gly Ala Ile Thr Ser Ser Asn Thr Ala Ala Thr  
 145 150 155 160  
 Asn Ala Ala Cys Ala Trp Leu Glu Ala Gln Glu Glu Glu Val Gly  
 165 170 175  
 Phe Pro Val Thr Pro Gln Val Pro Leu Arg Pro Met Thr Tyr Lys Ala  
 180 185 190  
 Ala Val Asp Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu Glu Gly  
 195 200 205  
 Leu Ile His Ser Gln Arg Arg Gln Asp Ile Leu Asp Leu Trp Ile Tyr  
 210 215 220  
 His Thr Gln Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro Gly Pro  
 225 230 235 240  
 Gly Val Arg Tyr Pro Leu Thr Phe Gly Trp Cys Tyr Lys Leu Val Pro  
 245 250 255  
 Val Glu Pro Asp Lys Val Glu Glu Ala Asn Lys Gly Glu Asn Thr Ser  
 260 265 270  
 Leu Leu His Pro Val Ser Leu His Gly Met Asp Asp Pro Glu Arg Glu  
 275 280 285  
 Val Leu Glu Trp Arg Phe Asp Ser Arg Leu Ala Phe His His Val Ala  
 290 295 300  
 Arg Glu Leu His Pro Glu Tyr Phe Lys Asn Cys Thr Ser Gly His His  
 305 310 315 320  
 His His His His

## (2) INFORMATION FOR SEQ ID NO:16:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1290 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

ATGGATCCAA	AAACTTTAGC	CCTTTCTTTA	TTAGCAGCTG	GCGTACTAGC	AGGTTGTAGC	60
AGCCATTCA	CAAATATGGC	GAATACCCAA	ATGAAATCG	ACAAATCAT	TATTGCTCAC	120
CGTGGTCTA	GCGTTTATT	ACCAAGAGCAT	ACGTTAGAAT	CTAAAGCACT	TGCGTTGCA	180
CAACAGGCTG	ATTATTTAGA	GCAAGATT	GCAATGACTA	AGGATGGTCG	TTTAGTGGTT	240
ATTACAGATC	ACTTTTTAGA	TGGCTTGACT	GATGTTGCGA	AAAAATTCCC	ACATCGTCAT	300
CGTAAAGATG	GCCGTTACTA	TGTCATCGAC	TTTACCTAA	AAGAAAATTCA	AAGTTTAGAA	360
ATGACAGAAA	ACTTTGAAAC	CATGGGTGGC	AAGTGGTCAA	AAAGTAGTGT	GGTTGGATGG	420
CCTACTGTAA	GGGAAAAGAAT	GAGACGAGCT	GAGCCAGCAG	CAGATGGGGT	GGGAGCAGCA	480
TCTCGAGACC	TGGAAAAACA	TGGAGCAATC	ACAAGTAGCA	ATACAGCAGC	TACCAATGCT	540
GCTTGTGCCT	GGCTAGAACG	ACAAGAGGAG	GAGGAGGTGG	GTTTCCAGT	CACACCTCAG	600
GTACCTTAA	GACCAATGAC	TTACAAGGC	GCTGTAGATC	TTAGCCACTT	TTTAAAAGAA	660
AAGGGGGGAC	TGGAAGGGCT	AATTCACTCC	CAACGAAGAC	AAGATATCCT	TGATCTGTGG	720
ATCTACCACA	CACAAGGCTA	CTTCCCTGAT	TGGCAGAACT	ACACACCAGG	GCCAGGGGTC	780
AGATATCCAC	TGACCTTTGG	ATGGTGTAC	AAGCTAGTAC	CAGTGAGCC	AGATAAGGTA	840
GAAGAGGCCA	ATAAAGGAGA	GAACACCAGC	TTGTTACACC	CTGTGAGCCT	GCATGGAATG	900

GATGACCCCTG	AGAGAGAAGT	GTTAGAGTGG	AGGTTTGACA	GCCGCCCTAGC	ATTCATCAC	960
GTGGCCCGAG	AGCTGCATCC	GGAGTACTTC	AAGAACTGCA	CTAGTGAGCC	AGTAGATCCT	1020
AGACTAGAGC	CCTGGAAGCA	TCCAGGAAGT	CAGCCTAAA	CTGCTTGTAC	CAATTGCTAT	1080
TGTAAGAAAGT	GTTGCTTCA	TTGCCAAGTT	TGTTTCATAA	CAAAGCCTT	AGGCATCTCC	1140
TATGGCAGGA	AGAACGGAG	ACAGCGACGA	AGACCTCCTC	AAGGCAGTCA	GACTCATCAA	1200
GTTTCTCTAT	CAAAGCAACC	CACCTCCAA	TCCCGAGGGG	ACCCGACAGG	CCCGAAGGAA	1260
ACTAGTGGCC	ACCATCACCA	TCACCATTA				1290

## (2) INFORMATION FOR SEQ ID NO:17:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 412 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Cys	Ser	Ser	His	Ser	Ser	Asn	Met	Ala	Asn	Thr	Gln	Met	Lys	Ser	Asp
1						5				10			15		
Lys	Ile	Ile	Ile	Ala	His	Arg	Gly	Ala	Ser	Gly	Tyr	Leu	Pro	Glu	His
						20			25			30			
Thr	Leu	Glu	Ser	Lys	Ala	Leu	Ala	Phe	Ala	Gln	Gln	Ala	Asp	Tyr	Leu
						35			40			45			
Glu	Gln	Asp	Leu	Ala	Met	Thr	Lys	Asp	Gly	Arg	Leu	Val	Val	Ile	His
					50			55			60				
Asp	His	Phe	Leu	Asp	Gly	Leu	Thr	Asp	Val	Ala	Lys	Lys	Phe	Pro	His
					65			70		75			80		
Arg	His	Arg	Lys	Asp	Gly	Arg	Tyr	Tyr	Val	Ile	Asp	Phe	Thr	Leu	Lys
					85			90			95				
Glu	Ile	Gln	Ser	Leu	Glu	Met	Thr	Glu	Asn	Phe	Glu	Thr	Met	Gly	Gly
					100			105			110				
Lys	Trp	Ser	Lys	Ser	Ser	Val	Val	Gly	Trp	Pro	Thr	Val	Arg	Glu	Arg
					115			120			125				
Met	Arg	Arg	Ala	Glu	Pro	Ala	Ala	Asp	Gly	Val	Gly	Ala	Ala	Ser	Arg
					130			135			140				
Asp	Leu	Glu	Lys	His	Gly	Ala	Ile	Thr	Ser	Ser	Asn	Thr	Ala	Ala	Thr
					145			150			155			160	
Asn	Ala	Ala	Cys	Ala	Trp	Leu	Glu	Ala	Gln	Glu	Glu	Glu	Val	Gly	
					165			170			175				
Phe	Pro	Val	Thr	Pro	Gln	Val	Pro	Leu	Arg	Pro	Met	Thr	Tyr	Lys	Ala
					180			185			190				
Ala	Val	Asp	Leu	Ser	His	Phe	Leu	Lys	Glu	Lys	Gly	Gly	Leu	Glu	Gly
					195			200			205				
Leu	Ile	His	Ser	Gln	Arg	Arg	Gln	Asp	Ile	Leu	Asp	Leu	Trp	Ile	Tyr
					210			215			220				
His	Thr	Gln	Gly	Tyr	Phe	Pro	Asp	Trp	Gln	Asn	Tyr	Thr	Pro	Gly	Pro
					225			230			235			240	
Gly	Val	Arg	Tyr	Pro	Leu	Thr	Phe	Gly	Trp	Cys	Tyr	Lys	Leu	Val	Pro
					245			250			255				
Val	Glu	Pro	Asp	Lys	Val	Glu	Glu	Ala	Asn	Lys	Gly	Glu	Asn	Thr	Ser
					260			265			270				
Leu	Leu	His	Pro	Val	Ser	Leu	His	Gly	Met	Asp	Asp	Pro	Glu	Arg	Glu
					275			280			285				
Val	Leu	Glu	Trp	Arg	Phe	Asp	Ser	Arg	Leu	Ala	Phe	His	His	Val	Ala
					290			295			300				

Arg	Glu	Leu	His	Pro	Glu	Tyr	Phe	Lys	Asn	Cys	Thr	Ser	Glu	Pro	Val
305				310				315						320	
Asp	Pro	Arg	Leu	Glu	Pro	Trp	Lys	His	Pro	Gly	Ser	Gln	Pro	Lys	Thr
					325				330					335	
Ala	Cys	Thr	Asn	Cys	Tyr	Cys	Lys	Lys	Cys	Cys	Phe	His	Cys	Gln	Val
					340			345					350		
Cys	Phe	Ile	Thr	Lys	Ala	Leu	Gly	Ile	Ser	Tyr	Gly	Arg	Lys	Lys	Arg
						355		360			365				
Arg	Gln	Arg	Arg	Arg	Pro	Pro	Gln	Gly	Ser	Gln	Thr	His	Gln	Val	Ser
						370		375			380				
Leu	Ser	Lys	Gln	Pro	Thr	Ser	Gln	Ser	Arg	Gly	Asp	Pro	Thr	Gly	Pro
					385		390		395				400		
Lys	Glu	Thr	Ser	Gly	His	His	His	His	His	His					
					405				410						

## (2) INFORMATION FOR SEQ ID NO:18:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 981 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

ATGGATCCAA	GCAGCCATT	ATCAAATATG	GC	GAATACCC	AA	ATGAAATC	AGACAAAATC	60
ATTATTGCTC	ACCGTGGTGC	TAGCGTTAT	TT	ACCAGAGC	AT	ACGTTAGA	ATCTAAAGCA	120
CTTGCCTTG	CACAACAGGC	TGATTATT	TA	GAGCAAGATT	TAG	CAATGAC	TAAGGTATGGT	180
CGTTTAGTGG	TTATTACACGA	TCACTTTT	TA	GATGGCTTGA	CTG	ATGTTGC	AAAAAAATT	240
CCACATCGTC	ATCGTAAAGA	TGGCCGTTAC	TAT	GTCATCG	ACT	TTACCTT	AAAAGAAATT	300
CAAAGTTTAG	AAATGACAGA	AAACTTTGAA	ACC	ATGGGTG	GCA	AGTGGTC	AAAAAGTAGT	360
GTGGTTGGAT	GGCCTACTGT	AAGGGAAAGA	ATG	GAGACGAG	CTG	AGCCAGC	AGCAGATGGG	420
GTGGGAGCAG	CATCTCGAGA	CCTGGAAAAA	CAT	GGAGCAGAA	TC	ACAAGTAG	CAATACAGCA	480
GCTACCAATG	CTGCTTGTC	CTGCTAGAA	GC	ACAAGAGG	AG	GAGGAGGT	GGGTTTTCCA	540
GTCACACCTC	AGGTACCTTT	AAGACCAATG	ACT	TACAAGG	CAG	CTGTAGA	TCTTAGCCAC	600
TTTTTAAAG	AAAAGGGGG	ACTGGAAGGG	CTA	ATTCACT	CCC	AAACGAAG	ACAAGATATC	660
CTTGATCTGT	GGATCTACCA	CACACAAGGC	TAC	TTCCCTG	ATT	GGCAGAA	CTACACACCA	720
GGGCCAGGGG	TCAGATATCC	ACTGACCTT	GG	ATGGTGCT	ACA	AGCTAGT	ACCAGTTGAG	780
CCAGATAAGG	TAGAAGAGGC	CAATAAAGGA	GAG	AACACCA	GCT	TGTTACA	CCCTGTGAGC	840
CTGCATGGAA	TGGATGACCC	TGAGAGAGAA	GTG	TTAGAGT	GGAGGTTGA	CAGCCGCC	900	
GCATTTCATC	ACGTGGCCCC	AGAGCTGCAT	CCGGAGTACT	TCAAGAACTG	CA	CTAGTGGC	960	
CACCATCAC	ATCACCATT	A						981

## (2) INFORMATION FOR SEQ ID NO:19:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 327 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Met	Asp	Pro	Ser	Ser	His	Ser	Ser	Asn	Met	Ala	Asn	Thr	Gln	Met	Lys
1									10					15	

Ser Asp Lys Ile Ile Ile Ala His Arg Gly Ala Ser Gly Tyr Leu Pro  
                   20                 25                 30  
 Glu His Thr Leu Glu Ser Lys Ala Leu Ala Phe Ala Gln Gln Ala Asp  
                   35                 40                 45  
 Tyr Leu Glu Gln Asp Leu Ala Met Thr Lys Asp Gly Arg Leu Val Val  
                   50                 55                 60  
 Ile His Asp His Phe Leu Asp Gly Leu Thr Asp Val Ala Lys Lys Phe  
                   65                 70                 75                 80  
 Pro His Arg His Arg Lys Asp Gly Arg Tyr Tyr Val Ile Asp Phe Thr  
                   85                 90                 95  
 Leu Lys Glu Ile Gln Ser Leu Glu Met Thr Glu Asn Phe Glu Thr Met  
                   100                105                110  
 Gly Gly Lys Trp Ser Lys Ser Ser Val Val Gly Trp Pro Thr Val Arg  
                   115                120                125  
 Glu Arg Met Arg Arg Ala Glu Pro Ala Ala Asp Gly Val Gly Ala Ala  
                   130                135                140  
 Ser Arg Asp Leu Glu Lys His Gly Ala Ile Thr Ser Ser Asn Thr Ala  
                   145                150                155                160  
 Ala Thr Asn Ala Ala Cys Ala Trp Leu Glu Ala Gln Glu Glu Glu  
                   165                170                175  
 Val Gly Phe Pro Val Thr Pro Gln Val Pro Leu Arg Pro Met Thr Tyr  
                   180                185                190  
 Lys Ala Ala Val Asp Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu  
                   195                200                205  
 Glu Gly Leu Ile His Ser Gln Arg Arg Gln Asp Ile Leu Asp Leu Trp  
                   210                215                220  
 Ile Tyr His Thr Gln Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro  
                   225                230                235                240  
 Gly Pro Gly Val Arg Tyr Pro Leu Thr Phe Gly Trp Cys Tyr Lys Leu  
                   245                250                255  
 Val Pro Val Glu Pro Asp Lys Val Glu Glu Ala Asn Lys Gly Glu Asn  
                   260                265                270  
 Thr Ser Leu Leu His Pro Val Ser Leu His Gly Met Asp Asp Pro Glu  
                   275                280                285  
 Arg Glu Val Leu Glu Trp Arg Phe Asp Ser Arg Leu Ala Phe His His  
                   290                295                300  
 Val Ala Arg Glu Leu His Pro Glu Tyr Phe Lys Asn Cys Thr Ser Gly  
                   305                310                315                320  
 His His His His His  
                   325

## (2) INFORMATION FOR SEQ ID NO:20:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1242 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

ATGGATCCAA GCAGCCATTG ATCAAATATG GCGAATACCC AAATGAAATC AGACAAAATC	60
ATTATTGCTC ACCGTGGTGC TAGCGGTTAT TTACCAGAGC ATACGTTAGA ATCTAAAGCA	120
CTTGCCTTG CACAACAGGC TGATTATTTA GAGCAAGATT TAGCAATGAC TAAGGATGGT	180
CGTTTAGTGG TTATTCACGA TCACTTTTTA GATGGCTTGA CTGATGTTGC GAAAAAATTC	240
CCACATCGTC ATCGTAAAGA TGGCCGTTAC TATGTCATCG ACTTTACCTT AAAAGAAATT	300

CAAAGTTAG	AAATGACAGA	AAACTTGTGAA	ACCATGGGTG	GCAAGTGGTC	AAAAAGTAGT	360
GTGGTTGGAT	GGCTTACTGT	AAGGGAAAGA	ATGAGACGAG	CTGAGCCAGC	AGCAGATGGG	420
GTGGGAGCG	CATCTCGAGA	CCTGGAAAAA	CATGGAGCAA	TCACAAGTAG	CAATAACAGCA	480
GCTACCAATG	CTGCTTGTGC	CTGGCTAGAA	GCACAAAGAGG	AGGAGGAGGT	GGGTTTTCCA	540
GTCACACCTC	AGGTACCTTT	AAGACCAATG	ACTTACAAGG	CAGCTGTAGA	TCTTAGCCAC	600
TTTTTAAAAG	AAAAGGGGGG	ACTGGAAGGG	CTAATTCACT	CCCAACGAAG	ACAAGATATC	660
CTTGATCTGT	GGATCTACCA	CACACAAAGG	TACTTCCCTG	ATTGGCAGAA	CTACACACCA	720
GGGCCAGGGG	TCAGATATCC	ACTGACCTT	GGATGGTGCT	ACAAGCTAGT	ACCAGTTGAG	780
CCAGATAAGG	TAGAAGAGGC	CAATAAAAGGA	GAGAACACCA	GCTTGTACAA	CCCTGTGAGC	840
CTGCATGGAA	TGGATGACCC	TGAGAGAGAA	GTGTTAGAGT	GGAGGGTTGA	CAGCCGCTA	900
GCATTTCATC	ACGTGGCCCC	AGAGCTGCAT	CCGGAGTACT	TCAAGAACTG	CACTAGTGAG	960
CCAGTAGATC	CTAGACTAGA	GCCCCTGGAAG	CATCCAGGAA	GTCAGCCTAA	AACTGCTTGT	1020
ACCAATTGCT	ATTGTAAGAAA	GTGTTGCTTT	CATTGCCAAG	TTTGTTCAT	AAACAAAAGCC	1080
TTAGGCATCT	CCTATGGCAG	GAAGAAGCGG	AGACAGCGAC	GAAGACCTCC	TCAAGGCAGT	1140
CAGACTCATC	AAGTTTCTCT	ATCAAAGCAA	CCCACCTCCC	AATCCCGAGG	GGACCCGACA	1200
GGCCCCGAAGG	AAACTAGTGG	CCACCATCAC	CATCACCAATT	AA		1242

(2) INFORMATION FOR SEQ ID NO:21:

**(i) SEQUENCE CHARACTERISTICS:**

- (A) LENGTH: 414 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Met	Asp	Pro	Ser	Ser	His	Ser	Ser	Asn	Met	Ala	Asn	Thr	Gln	Met	Lys
1					5				10					15	
Ser	Asp	Lys	Ile	Ile	Ile	Ala	His	Arg	Gly	Ala	Ser	Gly	Tyr	Leu	Pro
					20				25					30	
Glu	His	Thr	Leu	Glu	Ser	Lys	Ala	Leu	Ala	Phe	Ala	Gln	Gln	Ala	Asp
					35				40				45		
Tyr	Leu	Glu	Gln	Asp	Leu	Ala	Met	Thr	Lys	Asp	Gly	Arg	Leu	Val	Val
					50				55			60			
Ile	His	Asp	His	Phe	Leu	Asp	Gly	Leu	Thr	Asp	Val	Ala	Lys	Phe	
					65				70			75		80	
Pro	His	Arg	His	Arg	Lys	Asp	Gly	Arg	Tyr	Tyr	Val	Ile	Asp	Phe	Thr
					85				90			95			
Leu	Lys	Glu	Ile	Gln	Ser	Leu	Glu	Met	Thr	Glu	Asn	Phe	Glu	Thr	Met
					100				105			110			
Gly	Gly	Lys	Trp	Ser	Lys	Ser	Ser	Val	Val	Gly	Trp	Pro	Thr	Val	Arg
					115				120			125			
Glu	Arg	Met	Arg	Arg	Ala	Glu	Pro	Ala	Ala	Asp	Gly	Val	Gly	Ala	Ala
					130				135			140			
Ser	Arg	Asp	Leu	Glu	Lys	His	Gly	Ala	Ile	Thr	Ser	Ser	Asn	Thr	Ala
					145				150			155		160	
Ala	Thr	Asn	Ala	Ala	Cys	Ala	Trp	Leu	Glu	Ala	Gln	Glu	Glu	Glu	
					165				170			175			
Val	Gly	Phe	Pro	Val	Thr	Pro	Gln	Val	Pro	Leu	Arg	Pro	Met	Thr	Tyr
					180				185			190			
Lys	Ala	Ala	Val	Asp	Leu	Ser	His	Phe	Leu	Lys	Glu	Lys	Gly	Gly	Leu
					195				200			205			
Glu	Gly	Leu	Ile	His	Ser	Gln	Arg	Arg	Gln	Asp	Ile	Leu	Asp	Leu	Trp
					210				215			220			
Ile	Tyr	His	Thr	Gln	Gly	Tyr	Phe	Pro	Asp	Trp	Gln	Asn	Tyr	Thr	Pro

225	230	235	240
Gly Pro Gly Val Arg Tyr Pro Leu Thr Phe	Gly Trp Cys Tyr Lys	Leu	
245	250	255	
Val Pro Val Glu Pro Asp Lys Val Glu	Glu Ala Asn Lys	Gly Glu Asn	
260	265	270	
Thr Ser Leu Leu His Pro Val Ser Leu His	Gly Met Asp Asp Pro	Glu	
275	280	285	
Arg Glu Val Leu Glu Trp Arg Phe Asp Ser	Arg Leu Ala Phe His His		
290	295	300	
Val Ala Arg Glu Leu His Pro Glu Tyr Phe	Lys Asn Cys Thr Ser	Glu	
305	310	315	320
Pro Val Asp Pro Arg Leu Glu Pro Trp Lys	His Pro Gly Ser Gln Pro		
325	330	335	
Lys Thr Ala Cys Thr Asn Cys Tyr Cys Lys	Lys Cys Cys Phe His Cys		
340	345	350	
Gln Val Cys Phe Ile Thr Lys Ala Leu Gly	Ile Ser Tyr Gly Arg Lys		
355	360	365	
Lys Arg Arg Gln Arg Arg Pro Pro Gln Gly	Ser Gln Thr His Gln		
370	375	380	
Val Ser Leu Ser Lys Gln Pro Thr Ser Gln	Ser Arg Gly Asp Pro Thr		
385	390	395	400
Gly Pro Lys Glu Thr Ser Gly His His His	His His His His		
405	410		

## (2) INFORMATION FOR SEQ ID NO:22:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 288 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

ATGGAGCCAG TAGATCTTAG ACTAGAGCCC	TGGAAGCATC CAGGAAGTCA	GCCTAAAACT	60
GCTTGTACCA ATTGCTATTG TAAAAAGTGT	TGCTTTCATT GCCAAGTTTG	TTTCATAACA	120
GCTGCCCTTAG GCATCTCTA TGGCAGGAAG	AAGCGGAGAC AGCGACGAAG	ACCTCCTCAA	180
GGCAGTCAGA CTCATCAAGT TTCTCTATCA	AAGCAACCCA CCTCCCAATC	CAAAGGGGAG	240
CCGACAGGCC CGAAGGAAAC TAGTGGCCAC	CATCACCATC ACCATTAA		288

## (2) INFORMATION FOR SEQ ID NO:23:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 96 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Met Glu Pro Val Asp Pro Arg Leu Glu Pro	Trp Lys His Pro	Gly Ser	
1	5	10	15
Gln Pro Lys Thr Ala Cys Thr Asn Cys	Tyr Cys Lys Lys	Cys Cys Phe	
20	25	30	
His Cys Gln Val Cys Phe Ile Thr Ala	Ala Leu Gly Ile	Ser Tyr Gly	

35	40	45
Arg Lys Lys Arg Arg Gln Arg Arg Arg Pro Pro Gln Gly Ser Gln Thr		
50	55	60
His Gln Val Ser Leu Ser Lys Gln Pro Thr Ser Gln Ser Lys Gly Glu		
65	70	75
Pro Thr Gly Pro Lys Glu Thr Ser Gly His His His His His His		
85	90	95

## (2) INFORMATION FOR SEQ ID NO:24:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 909 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

ATGGGTGGCA AGTGGTCAAA AAGTAGTGTG GTTGGATGGC CTACTGTAAG GGAAAGAACAT	60
AGACGAGCTG AGCCAGCAGC AGATGGGGTG GGAGCAGCAT CTCGAGACCT GGAAAAACAT	120
GGAGCAATCA CAAGTAGCAA TACAGCAGCT ACCAATGCTG CTTGTGCCTG GCTAGAAGCA	180
CAAGAGGAGG AGGAGGTGGG TTTTCCAGTC ACACCTCAGG TACCTTTAAG ACCAATGACT	240
TACAAGGCAG CTGTAGATCT TAGCCACTTT TTAAAAGAAA AGGGGGACT GGAAGGGCTA	300
ATTCACTCCC AACGAAGACA AGATATCCTT GATCTGTGGA TCTACCACAC ACAAGGCTAC	360
TTCCCTGATT GGCAGAACTA CACACCAGGG CCAGGGGTCA GATATCCACT GACCTTTGGA	420
TGGTGCTACA AGCTAGTACC AGTTGAGCCA GATAAGGTAG AAGAGGCCAA TAAAGGAGAG	480
AACACCAGCT TGTTACACCC TGTGAGCCTG CATGGAATGG ATGACCTCTGA GAGAGAAGTG	540
TTAGAGTGGG GGTTTGACAG CGCGCTAGCA TTTCATCACG TGGCCCGAGA GCTGCATCCG	600
GAGTACTTCA AGAACTGCAC TAGTGAGCCA GTAGATCCTA GACTAGAGCC CTGGAAAGCAT	660
CCAGGAAGTC AGCCTAAAC TGCTTGCTAC AATTGCTATT GTAAAAAGTG TTGCTTTCAT	720
TGCCAAGTTT GTTTCATAAC AGCTGCCTA GGCATCTCCT ATGGCAGGAA GAAGCGGGAGA	780
CAGCGACGAA GACCTCCTCA AGGCAGTCAG ACTCATCAAG TTTCTCTATC AAAGCAACCC	840
ACCTCCCCAT CCAAAGGGGA GCCGACAGGC CCGAAGGAAA CTAGTGGCCA CCATCACCAT	900
CACCATTA	909

## (2) INFORMATION FOR SEQ ID NO:25:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 303 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

Met Gly Gly Lys Trp Ser Lys Ser Ser Val Val Gly Trp Pro Thr Val			
1	5	10	15
Arg Glu Arg Met Arg Arg Ala Glu Pro Ala Ala Asp Gly Val Gly Ala			
20	25	30	
Ala Ser Arg Asp Leu Glu Lys His Gly Ala Ile Thr Ser Ser Asn Thr			
35	40	45	
Ala Ala Thr Asn Ala Ala Cys Ala Trp Leu Glu Ala Gln Glu Glu Glu			
50	55	60	
Glu Val Gly Phe Pro Val Thr Pro Gln Val Pro Leu Arg Pro Met Thr			
65	70	75	80

15 / 15

Tyr	Lys	Ala	Ala	Val	Asp	Leu	Ser	His	Phe	Leu	Lys	Glu	Lys	Gly	
					85					90				95	
Leu	Glu	Gly	Leu	Ile	His	Ser	Gln	Arg	Arg	Gln	Asp	Ile	Leu	Asp	Leu
					100					105				110	
Trp	Ile	Tyr	His	Thr	Gln	Gly	Tyr	Phe	Pro	Asp	Trp	Gln	Asn	Tyr	Thr
					115					120				125	
Pro	Gly	Pro	Gly	Val	Arg	Tyr	Pro	Leu	Thr	Phe	Gly	Trp	Cys	Tyr	Lys
					130					135				140	
Leu	Val	Pro	Val	Glu	Pro	Asp	Lys	Val	Glu	Glu	Ala	Asn	Lys	Gly	Glu
					145					150				160	
Asn	Thr	Ser	Leu	Leu	His	Pro	Val	Ser	Leu	His	Gly	Met	Asp	Asp	Pro
					165					170				175	
Glu	Arg	Glu	Val	Leu	Glu	Trp	Arg	Phe	Asp	Ser	Arg	Leu	Ala	Phe	His
					180					185				190	
His	Val	Ala	Arg	Glu	Leu	His	Pro	Glu	Tyr	Phe	Lys	Asn	Cys	Thr	Ser
					195					200				205	
Glu	Pro	Val	Asp	Pro	Arg	Leu	Glu	Pro	Trp	Lys	His	Pro	Gly	Ser	Gln
					210					215				220	
Pro	Lys	Thr	Ala	Cys	Thr	Asn	Cys	Tyr	Cys	Lys	Lys	Cys	Cys	Phe	His
					225					230				235	
Cys	Gln	Val	Cys	Phe	Ile	Thr	Ala	Ala	Leu	Gly	Ile	Ser	Tyr	Gly	Arg
					245					250				255	
Lys	Lys	Arg	Arg	Gln	Arg	Arg	Arg	Pro	Pro	Gln	Gly	Ser	Gln	Thr	His
					260					265				270.	
Gln	Val	Ser	Leu	Ser	Lys	Gln	Pro	Thr	Ser	Gln	Ser	Lys	Gly	Glu	Pro
					275					280				285	
Thr	Gly	Pro	Lys	Glu	Thr	Ser	Gly	His	His	His	His	His	His		
					290					295				300	

**(2) INFORMATION FOR SEQ ID NO:26:**

### (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 57 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

**TTCGAACCA TGGCCGCGGA CTAGGGCCA CCATCACCAT CACCATTAAC GGAATT**

57

(2) INFORMATION FOR SEQ ID NO:27:

**(i) SEQUENCE CHARACTERISTICS:**

- (A) LENGTH: 17 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

**Thr Ser Gly His His His His His**

1 - 5

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 98/06040

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/49 C12N15/62 C07K14/16 A61K39/21

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the International search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 94 04686 A (BARSOUM JAMES G ;BIOGEN INC (US); FAWELL STEPHEN E (US); PEPINSKY) 3 March 1994 see page 54 - page 73	1,4, 13-15
X	BODÉUS M ET AL.: "In vitro binding and phosphorylation of human immunodeficiency virus type 1 Nef protein by serine/threonine protein kinase" JOURNAL OF GENERAL VIROLOGY, vol. 76, no. 6, June 1995, pages 1337-1344, XP002092508. READING GB see page 1338, left-hand column, paragraph 3	1,5, 13-15

 Further documents are listed in the continuation of box C. Patent family members are listed in annex.

## \* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the International search

5 February 1999

Date of mailing of the International search report

18/02/1999

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentstaan 2  
 NL - 2280 HV Rijswijk  
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
 Fax (+31-70) 340-3016

Authorized officer

Cupido, M

## INTERNATIONAL SEARCH REPORT

Inte...onal Application No  
PCT/EP 98/06040

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SALFELD J ET AL: "A tripartite HIV-1 tat-env-rev fusion protein" EMBO JOURNAL, vol. 9, no. 3, 1 March 1990, pages 965-970, XP000113784 see the whole document	1,4
X	AHMED A AZAD ET AL: "Large-scale production and characterization of recombinant human immunodeficiency virus type 1 Nef" JOURNAL OF GENERAL VIROLOGY, vol. 75, no. 3, 1 March 1994, pages 651-655, XP000565729 see the whole document	1,5, 13-15
A	JANSON H ET AL.: "Protein D, the immunoglobulin D-binding protein of Haemophilus influenzae, is a lipoprotein" INFECTION AND IMMUNITY, vol. 60, no. 4, April 1992, pages 1336-1342, XP002092509 WASHINGTON US cited in the application see the whole document	6-8

**INTERNATIONAL SEARCH REPORT**

Information on patent family members

Inter. Application No.

PCT/EP 98/06040

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 9404686	A 03-03-1994	AT	173016 T	15-11-1998
		AU	667244 B	14-03-1996
		AU	5083293 A	15-03-1994
		CA	2135642 A	03-03-1994
		DE	69321962 D	10-12-1998
		DE	656950 T	14-03-1996
		EP	0656950 A	14-06-1995
		ES	2123062 T	01-01-1999
		FI	945248 A	05-01-1995
		JP	10033186 A	10-02-1998
		JP	2702285 B	21-01-1998
		JP	7503617 T	20-04-1995
		NO	944273 A	17-02-1995
		NZ	255831 A	24-04-1997
		US	5674980 A	07-10-1997
		US	5670617 A	23-09-1997
		US	5652122 A	29-07-1997
		US	5747641 A	05-05-1998